# Functional Fertility Genomics in Sheep

**Kisun Pokharel**

Natural Resources Institute Finland (Luke)
And
Doctoral Programme in Integrative Life Science (ILS)
Faculty of Biological and Environmental Sciences
University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination with the permission
of the Faculty of Biological and Environmental Sciences of the
University of Helsinki, in Auditorium 2041, Biocenter 2, Helsinki
on the $24^{th}$ of November, 2020 at 12 noon.

Helsinki 2020

**Supervisors**
Professor Juha Kantanen
Natural Resources Institute Finland (Luke), Finland

Professor Meng-Hua Li
China Agricultural University, China

**Pre-examiners**
Dr. Stéphane Fabre
French National Institute for Agricultural Research INRA, France

Professor Thomas E. Spencer
University of Missouri, USA

**Opponent**
Professor David MacHugh
University College Dublin, Ireland

**Custos**
Professor Craig Primmer
University of Helsinki, Finland

**Contact information**
Kisun Pokharel
Natural Resources Institute Finland (Luke)
FI-31600 Jokioinen, Finland
Email: kisun.pokharel@luke.fi

Cover picture: ©ProAgria Southern Finland (photo by Ilkka Valkila)

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis 69/2020
ISSN 2342-3161 (paperback)
ISSN 2341-317X (PDF)
ISBN 978-951-51-6674-6 (paperback)
ISBN 978-951-51-6675-3 (PDF)
http://ethesis.helsinki.fi

Unigrafia Oy
Helsinki 2020

***

# Table of Contents

# List of Original Publications

This thesis is based on the following publications:

I Pokharel K, Peippo J, Andersson G, Li M-H & Kantanen J. 2015. Transcriptome profiling of Finnsheep ovaries during out-of-season breeding period. *Agricultural and Food Science* 24: 1-9. DOI: 10.23986/afsci.46512

II Hu X-J, Pokharel K, Peippo J, Ghanem N, Zhaboyev I, Kantanen J & Li M-H. 2016. Identification and characterization of miRNAs in the ovaries of a highly prolific sheep breed. *Animal Genetics* 47: 234-239. DOI: 10.1111/age.12385

III Pokharel K, Peippo J, Honkatukia M, Seppälä A, Rautiainen J, Ghanem N, Hamama T-M, Crowe M, Andersson M, Li M-H & Kantanen J. 2018. Integrated ovarian mRNA and miRNA transcriptome profiling characterizes the genetic basis of prolificacy traits in sheep (*Ovis aries*). *BMC Genomics* 19: 104. DOI: 10.1186/s12864-017-4400-4

IV Pokharel K, Peippo J, Weldenegodguad M, Honkatukia M, Li M-H & Kantanen J. 2020. Gene expression profiling of *corpus luteum* reveals important insights about early pregnancy in domestic sheep. *Genes* 11: 415. DOI: 10.3390/genes11040415

V Pokharel K, Peippo J, Li M-H & Kantanen J. 2020. Functional profiling of endometrium transcriptome during preimplantation development in Finnsheep, Texel and their F1 crosses. *Agricultural and Food Science* 29: 331-345. DOI: 10.23986/afsci.90523

VI Pokharel K, Peippo J, Li M-H & Kantanen J. 2020. Identification and characterization of miRNAs during early pregnancy in domestic sheep. *Animal Genetics* 51: 833-836. DOI: 10.1111/age.12992

VII Yang J*, Li X*, Cao Y-H*, Pokharel K*, Hu X-J*, Chen Z-H*, Xu S-S, Peippo J, Honkatukia M, Kantanen J & Li M-H. 2019. Comparative mRNA and miRNA expression in European mouflon (*Ovis musimon*) and sheep (*Ovis aries*) provides novel insights into the genetic mechanisms for female reproductive success. *Heredity* 122: 172-186. DOI: 10.1038/s41437-018-0090-1
*These authors contributed equally

The publications are referred to in the text by their roman numerals. All publications except II are open access articles published under Creative Commons licenses. The paperback version of this thesis includes PDF copies of the original articles.

## Author's contributions to publications

I Data analysis, interpretation and manuscript writing.

II Contribution to data analysis and manuscript writing.

III–VI Data analysis, interpretation and manuscript writing.

VII Contribution to data analysis.

# List of abbreviations

| **Abbreviation** | **Definition** |
|---|---|
| BCS | Body condition score |
| bp | Base pairs |
| C18MC | Chromosome 18 miRNA cluster |
| *CA5A* | Carbonic anhydrase VA |
| CL | *Corpus luteum* |
| DE | Differential expression (analysis) |
| DEG | Differentially expressed gene |
| ERV | Endogenous retrovirus |
| *GDF9* | Growth differentiation factor 9 |
| GnRH | Gonadotrophin releasing hormone |
| GO | Gene Ontology |
| IFNT | Interferon tau |
| JSRV | Jaagsiekte sheep retrovirus |
| Kb | Kilobases |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| LFC | $Log_2$(fold-change) |
| LH | Luteinizing hormone |
| miRNA | MicroRNA |
| mRNA | Messenger RNA |
| NR | Non-redundant |
| P4 | Progesterone |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| POE | Parent-of-origin effect |
| qRT-PCR | quantitative reverse transcription-PCR |
| RNA-seq | RNA sequencing |
| SNP | Single-nucleotide polymorphism |
| TGF-$\beta$ | Transforming growth factor beta |
| *vs.* | versus |

# Abstract

The conservation, characterization, and sustainable utilization of farm animal genetic resources have far-reaching socioeconomic effects. Finnsheep, which is native to Finland, is globally known for its exceptional prolificacy and has been exported to several countries to improve the productivity of local breeds. Prolificacy, mainly determined by the ovulation rate and litter size phenotypes, has high economic importance in sheep farming systems. The ovulation rate is associated with the number of ovulatory follicles produced during the follicular growth phase of the estrous cycle, and successful implantation of the conceptus ensures maximum litters.

The primary aim of this study was to evaluate the exceptional prolificacy of Finnsheep. Another important and overarching aim was to explore the genetic aspects of two important reproductive events leading to ovulation and successful pregnancy in sheep. To gain an insight into how these two processes are regulated genetically and non-genetically, the whole transcriptomes of selected reproductive tissues relevant to folliculogenesis (first phase, ovary) and early pregnancy prior to implantation (second phase, *corpus luteum (CL)* and endometrium) were analyzed. In the main experiment, the ovaries of 31 Finnsheep (n = 11) and Texel (n = 11) ewes as well as their F1 crosses (n = 9) collected during surgery were biopsied; other tissues, including the CL and endometrium, were collected at the slaughterhouse. Half of the animals were maintained on a flushing diet during the entire experiment to study the influence of nutrition during early pregnancy. In addition, the transcriptomes of the ovarian and endometrial tissues of Finnsheep ewes were compared to those of European feral mouflons. For transcriptomic analysis, cDNA libraries targeting mRNAs and microRNAs were sequenced using the Illumina HiSeq 2000 system with 100-base pair (bp) paired-end and 50-bp single-end techniques, respectively.

Gene expression results from the first phase of the experiment revealed that the flushing diet strongly affected the Texel and, to some extent, the F1-crosses of sheep; however, Finnsheep were not affected by the diet. Likewise, the gene expression profiles of the F1 crosses were more similar to those of Finnsheep than of Texel, possibly because of the parent-of-origin effects, such as genomic imprinting. While no major candidate genes reportedly associated with prolificacy were differ-

entially or highly expressed, FecG$^F$ polymorphism (V371M) in *GDF9* was present in half of the experimental Finnsheep and F1 crosses but completely absent in the Texel sheep.

The second phase of the experiment provided an insight on the cross-talk between the CL and endometrium, as revealed by both the shared and tissue-specific genes. Finnsheep had a higher embryo mortality rate than Texel. The top expressed genes were associated with progesterone formation and events (elongation and attachment), leading to blastocyst implantation. Several endogenous retroviruses (ERVs) were expressed in the endometrium samples. Immune-related genes and pathways appeared to be differentially expressed and/or regulated in the CL of pure-breds. During the preimplantation stage, the immune system of the Finnsheep ewes appeared superior to that of the Texel ewes. Owing to the influence of interferon tau in massive gene expression changes starting around day 13, the gene expression profiles of the endometrium samples in response to elongated embryos were remarkably different from those in response to spherical embryos. Moreover, several interferon-stimulated genes were upregulated in the endometrium samples having elongated embryos. In addition to differentially expressed genes between the reproductive tissues of domestic (Finnsheep) and feral (European mouflon) sheep, we identified novel genes and miRNAs and their regulatory mechanisms associated with reproductive traits.

A few novel structural findings were obtained as part of this work. Maternally imprinted miRNA cluster on chromosome 18 (C18MC), which is conserved exclusively among placental mammals, was reported for the first time in sheep. With 46 expressed miRNAs, sheep C18MC may be the largest miRNA cluster among all mammalian species. The miRNAs (>500) quantified in this thesis are valuable, given that only 153 sheep miRNAs have been identified till date. A novel ERV transcript with high similarity to the genomic region within the FecL locus and known to affect prolificacy was identified, indicating that ERV plays an important role in reproduction and may even contribute to litter size differences.

Overall, the work presented in this thesis explored two critically important aspects of reproduction, i.e., ovulation and preimplantation, in sheep using state-of-the-art genomics and bioinformatics tools. The resources and findings from this thesis are highly relevant to both breeders and researchers studying sheep. The comprehensive list of genes and miRNAs expressed in the three key reproductive tissues is a useful resource for understanding transcriptional patterns during the follicular growth phase and preimplantation stage leading to pregnancy in sheep. The diet- and F1 cross-related experimental results will be valuable for implementing sheep breeding strategies aimed at achieving an optimum reproductive capacity. Furthermore, these results will provide a foundation for future research, and the data may have additional applications following the advancement of analysis tools and technologies.

# Chapter 1

# Introduction

## 1.1 Background and Significance

Sheep were originally domesticated in the fertile crescent during the prehistoric period (11,000 – 9,000 years before present) (Ryder, 1984). Since then, as one of the first domesticated species, sheep have constituted an important part of the global agricultural economy. Although they were originally reared for meat, sheep were eventually selected for their secondary products such as milk and wool in southwest Asia and Europe approximately 5,000 and 4,000 years ago, respectively (Chessa et al., 2009; Pedrosa et al., 2005). Because of the preferential selection of sheep for their secondary products, many primitive domestic breeds were lost, although some survived by returning to feral or semi-feral conditions (Sherratt, 1981). Breeds such as Finnsheep, Romanov, and Icelandic sheep are considered native pure-breeds in Europe and have been known to exist since the Iron Age (Ryder, 1981). During the $20^{th}$ century, increases in selection and intensive breeding led to the emergence of modern breeds; currently, the number of sheep breeds (>1,500) is greater than that of any other livestock species, which have adapted to diverse climatic conditions[1]. Sheep are multipurpose livestock species; they produce wool and yield meat. They are also raised for their pelts, as dairy animals, and as model animals for scientific research. Sheep are used for landscape management practices and are socioeconomically important.

Because of the varying litter size phenotype, sheep reproduction has been extensively researched. Unlike other domestic species that typically produce 1–2 (cattle and goats) or ≥4 offspring (pigs and dogs), sheep have a larger litter size (1–8) variation (Fabre et al., 2006). Prolific sheep breeds are valuable to the global sheep industry as genetic resources, and their prolific traits, including the high ovulation rate and large litter size, affect both the biological and financial performances of

[1]The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture (http://www.fao.org/3/a-i5077e.pdf)

sheep production systems. However, traditional breeding approaches are often ineffective because of the low heritability and sex-limited expression of prolificacy traits. The complexity of prolificacy traits with multiple genetic, epistatic, and environmental determinants represents a significant challenge in genetic analyses because of difficulties in isolating the phenotype of one gene amid the many other genetic and environmental effects. Thus, studying one gene at a time would not be feasible; studies of the genetics of prolificacy in sheep will require an analysis of the interactions among, and environmental dependency of, genetic effects by integrating various technologies and datasets to form a unified concept. Therefore, in this thesis, structural and functional changes in gene expression during the reproductive cycle in different breeds were investigated to identify the key regulators of this complex trait.

The "genomics era" has been characterized by the fusion of computer science, molecular biology, and engineering in the last decade, and this has revolutionized the way research is conducted. Modern genomic tools and methods provide high-resolution data that can be utilized to gain an insight into how such complex traits are regulated at the basic level of DNA and RNA. Research on sheep has been greatly facilitated by advancements in genomic methods and the availability of Illumina ovine (50k and more recently 600k) single-nucleotide polymorphism (SNP) beadchips, as well as a fully annotated sheep reference genome (Jiang et al., 2014), both of which have been made available as part of the International Sheep Genomics Consortium (ISGC). Application of high-throughput technologies such as SNP genotyping (e.g., Demars et al. (2013); Xu et al. (2018)), whole-genome sequencing (e.g., Nosrati et al. (2019); Wang et al. (2019)), whole-genome bisulfite sequencing (e.g., Zhang et al. (2017)), and RNA sequencing (RNA-seq) have yielded prolificacy-related genetic markers and candidate genes for several breeds. Transcriptome-based studies are typically performed (see Table 1.2) to identify and characterize genes associated with economic traits (e.g., meat quality, wool, and prolificacy) and other medical purposes.

In this thesis, RNA-seq was used to explore the whole transcriptome profiles of selected reproductive tissues collected during two phases of the reproductive cycle in sheep (see 3.2 for the experimental design). The primary goal was to study the exceptional prolificacy of Finnsheep: **What makes Finnsheep possibly the most prolific sheep breed in the world?** The hypothesis was that a certain genetic makeup and environmental factors are key to this exceptional phenotype. To support this hypothesis, Finnsheep were compared with another sheep breed—the Texel breed—with a high global importance but low prolificacy. The average litter size of Texel sheep is approximately half (1.5) of that of Finnsheep (2.8). Furthermore, including cross-breeds can provide an insight into the heritability of genetic markers associated with the phenotype. To understand the role of diet, often considered as one of the most important environmental parameters leading to differences in the ovulation rate, in this phenotype, half of the ewes

were maintained on a flushing diet throughout the experiment. Additionally, to gain an insight into the evolution of sheep reproduction, tissue samples from feral European mouflon (*Ovis orientalis musimon*) were studied together with those from Finnsheep.

## 1.2 Physiology of female reproduction in sheep

### 1.2.1 Estrous cycle

Reproduction is a fundamental process in the evolution of all living organisms on the earth. The estrous cycle is the basis of reproduction in sheep and other seasonal breeders. Sheep are seasonally polyestrous, and each estrous cycle lasts 13–19 days, the average being 17 days. Additionally, sheep are short-day breeders, indicating that autumn is their preferred breeding season. However, some breeds have extended breeding seasons (e.g., Merino and Dorset), and others may breed throughout the year (e.g., Finnsheep, Romanov, and Hu). The estrous cycle is divided into four phases: proestrus, estrus, metestrus, and diestrus. The estrus phase, which lasts for 24–36 h, is characterized as the time during which ewes are in heat and ready for mating. The estrous cycle is regulated by hormones and other signals produced in the hypothalamus, pituitary gland, *corpus luteum* (CL), and uterus. The hypothalamus secretes gonadotrophin-releasing hormone (GnRH); the anterior pituitary gland releases luteinizing hormone (LH), follicle-stimulating hormone (FSH), and oxytocin; the CL produces progesterone (P4) and oxytocin; and the endometrium secretes prostaglandin alpha 2-alpha (PGF2$\alpha$) (reviewed in Scaramuzzi et al. (1993)). LH induces ovulation, aids in follicle maturation, and supports CL development. Similarly, FSH induces follicle development and estrogen production (Vegetti and Alagna, 2006).

### 1.2.2 Ovarian folliculogenesis

Ovaries have two important functions: the production of oocytes and secretion of hormones essential for fertilization, and pregnancy establishment in the reproductive tract. Follicles, the basic functional units of ovaries, are comprised of oocytes surrounded by multiple layers of somatic cells. Throughout the active reproductive lifespan of a female, during each reproductive cycle, a small subset of primordial follicles from the pre-established ovarian reserve is recruited to the growing pool of primary and small preantral follicles, often referred to as the dynamic ovarian reserve (Monniaux et al., 2014). Gonadotrophin-independent preantral follicle development involves cross-talk between the oocyte and its surrounding cells, which secrete several oocyte-derived growth factors including members of the transforming growth factor $\beta$ (TGF-$\beta$) superfamily. During further development, an antrum cavity is formed, and the follicles are referred to as antral follicles. FSH plays an

important role during transition from the pre-antral to antral stage. The next stage of development is mediated by pituitary gonadotrophins, FSH, and LH. The final stage of antral follicle development proceeds in a wave-like pattern in which 3–4 waves each, lasting for 4–6 days, occur in ewes (Fig. 1.1, Ginther et al. (1995)). Finnsheep ewes reportedly have four waves starting on days 1, 6, 10, and 13 on average (Bartlewski et al., 1999). An increased FSH concentration precedes each follicular wave, from which one (mono-ovulatory) or more (polyovulatory) dominant follicles continue to grow, whereas all others regress through atresia. Particularly in prolific ewes, follicles from two consecutive follicular waves can ovulate simultaneously (Bartlewski et al., 2011). During this phase of dominant follicle selection, rapid regression of P4 occurs. Consequently, low P4 concentrations during the follicular phase allow for an increase in the GnRH pulse frequency, thereby increasing estradiol production by the dominant follicle and eliciting estrus, an LH surge, and ovulation (Sarda et al., 1973).

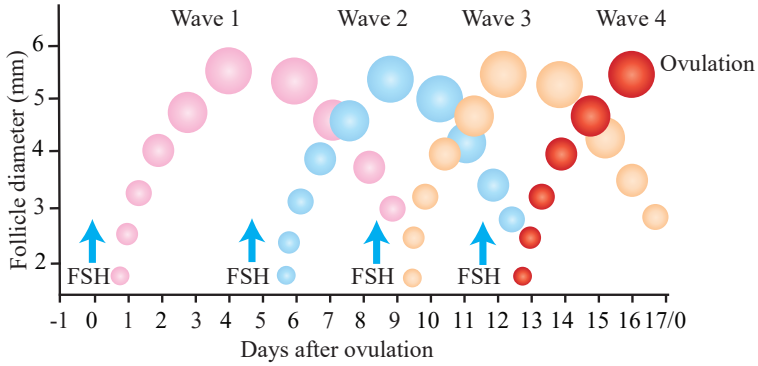


Figure 1.1: **Wave-like pattern of follicular growth in sheep**. Schematic illustration showing four follicular waves, of which typically the first three are non-ovulatory and the fourth yields ovulation of dominant follicle(s). In some breeds such as prolific Finnsheep, dominant follicles from two consecutive follicular waves, wave 3 and 4 in this case, have been found to ovulate simultaneously (Bartlewski et al., 2011). A peak in FSH (blue arrows) precedes each wave leading to the emergence of gonadotrophin-dependent follicles

.

The optimal growth and selection of dominant follicles are linked to the ovulation rate and are thus critical to sheep production systems. A hypothetical model involving FSH level manipulation to control the ovulation rate has been in place for nearly 35 years (Baird, 1987; Scaramuzzi et al., 2011). Additionally, the model, referred to as the "FSH window theory," suggests that the ovulation rate can be increased either by maintaining the FSH concentration above a threshold for a longer time or by increasing the number of ovulatory follicles (see Fig. 5 in Scaramuzzi et al. (2011)). Mutations in several candidate genes (see also 1.3.1), particularly those of the TGF-$\beta$ family, reportedly influence folliculogenesis, and

consequently, the ovulation rate (Monniaux et al., 2014).

### 1.2.3 Establishment of pregnancy in sheep

The fertilization of an ovulated follicle-derived oocyte results in the formation of a single binucleated cell known as the zygote, which undergoes a series of mitotic divisions leading to the morula stage (24–40 cells) of embryo formation; the formed embryo enters the uterus between days 4 and 6 post mating (Fig. 1.2). By day 7, the blastocyst is formed and is characterized by two distinct cell populations: the inner cell mass and the trophectoderm or extraembryonic tissue (Spencer and Hansen, 2015). The blastocyst stage marks the beginning of pregnancy establishment in domestic ruminants and involves the maternal recognition of pregnancy, implantation, and placentation (Spencer et al., 2004a).

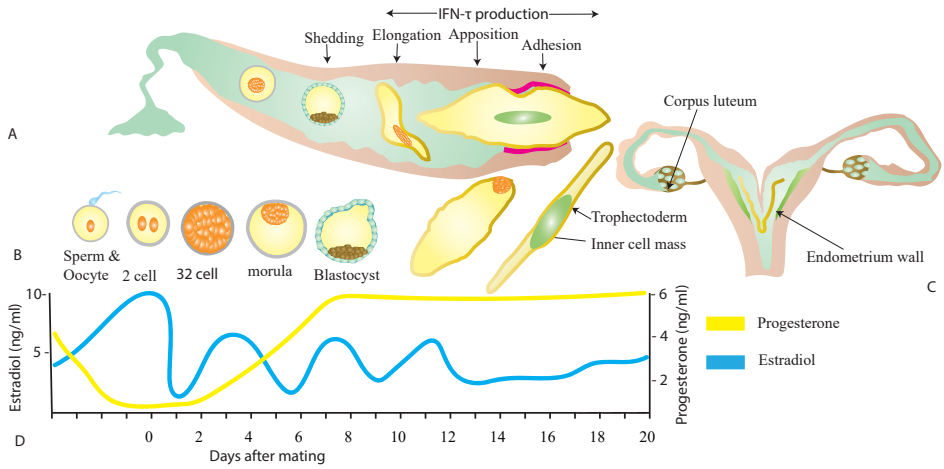


Figure 1.2: **Overview of pregnancy establishment in sheep**. (A) Different stages of implantation in uterus, (B) different stages of blastocyst growth, from zygote formation to conceptus development, (C) the uterus showing the implanted conceptus, and (D) the progesterone and estradiol levels during preimplantation. Figure adapted from Spencer et al. (2007)

Maternal recognition of pregnancy (occurs at approximately days 12 and 13 in sheep), an important event during preimplantation, is mediated by an antiluteolytic substance known as interferon tau (IFNT); IFNT is released from the mononuclear trophectoderm of an elongated conceptus (Bazer, 2013; Brooks et al., 2014; Brooks and Spencer, 2015) and is produced between days 10 and 21 and peaks on days 12 to 16 (Farin et al., 1990). In addition to IFNT, an elongated conceptus also secretes prostaglandins and cortisol. The role of P4, IFNT, and other factors during early pregnancy in sheep have been studied by other researchers (Brooks et al., 2014; Brooks, 2016; Spencer et al., 2016). Implantation in sheep involves the adhesion of mononucleated cells of the trophectoderm to the luminal

epithelia (LE) of the endometrium. The key stages are: (1) shedding of the zona pellucida, (2) conceptus elongation, (3) apposition, and (4) adhesion (Fig. 1.2). Shedding of the zona pellucida occurs by day 9; after this, the blastocyst undergoes a series of structural changes from spherical to tubular (days 12–13), leading to an elongated filamentous conceptus (between days 13 and 19). Conceptus elongation marks the beginning of implantation and involves apposition (days 12–16) and firm adhesion (by day 16) to the LE of the endometrium.

Several studies (e.g., Bolet (1986); Quinlivan et al. (1966)) have shown that most cases of embryonic death occur during the preimplantation stage. The CL and endometrium are the two vital players for successful pregnancy establishment and are thus critical for determining litter sizes and reproductive success. The CL is a temporary organ that forms from the ovulatory follicle and degenerates or regresses after implantation; moreover, it produces P4, which is essential for maintaining the uterine environment during early pregnancy (Spencer et al., 2004b). The CL is composed of four different cell types, including two (small and large) steroidogenic luteal cells, capillary endothelial cells, and fibroblasts (Farin et al., 1989). In addition to being the main site for blastocyst implantation, the LE of the endometrium secretes histotroph, which is essential for pregnancy establishment and placentation (Spencer et al., 2008; Filant and Spencer, 2014).

## 1.3 Sheep prolificacy

Two important fertility traits—the ovulation rate and litter size—confer a high economic value to sheep and influence the efficiency and profitability of sheep farming systems (Notter, 2008). In sheep and other multiparous mammals, prolificacy is determined by the ovulation rate and litter size. Litter size is an important parameter in sheep production systems. However, low heritability and various external factors, such as feed availability, geographic and climatic conditions, and maternal care, are attributable to the litter size difference in sheep. In fact, the litter size variation in sheep is higher than that in any other domestic animal. Therefore, the optimal litter size may not always be the largest possible and is important in different production systems. Generally, a lamb mortality rate below or above 15–20% can be used as a reference to target the optimal litter size. A large litter size is not desirable when the mortality rate of lambs is greater than 15–20%, and a mortality rate lower than this indicates the potential for improving overall productivity. Depending on the situation, the optimal litter size can range from one lamb to quadruplets. For example, in household farms in a mountainous region in Nepal, where the soil is not fertile and the climate is not favorable, the optimal litter size target would be one lamb. Similarly, a household in a hilly region may target duplets because of the better conditions. In contrast, a commercial farming system with a proper management system, sufficient fodder, good maternal care,

and supervised lambing with fostering may target even larger litter sizes.

### 1.3.1 Genetic factors affecting sheep prolificacy

The ovulation rate and litter size of sheep breeds can be regulated by a combination of small individual effects of different genes (Ricordeau et al., 1990) or by a single genetic marker with major effects, known as fecundity genes (Chu et al., 2007; Davis, 2005). Several mutations and other polymorphisms in *GDF9*, *BMP15*, and *BMPR1B*—three major genes from the TGF-$\beta$ family—and more recently in *B4GALNT2* are known to affect the ovulation rate and litter size of different sheep breeds (Table 1.1). The physiological mechanisms underlying the role of these mutations on the ovulation rate are still unclear. Additionally, parent-of-origin effects, such as imprinting and epigenetic markers, are associated with prolificacy (Coster et al., 2012; Davis et al., 2001).

Table 1.1: **Causative mutations of four major genes (*BMP15*, *BMPR1B*, *GDF9*, and *B4GALNT2*) in different sheep breeds.** Note1* - deletion of amino acid fragment P154-S159; Note2* - a polymorphism that includes a single-nucleotide substitution (G301T), a 3-bp deletion (c.302-304delCTA), and a C insertion (c.310insC); Note3* - flock of Brazilian breeds including only those ewes with at least one triplet birth history; Note4* - FecL is a 194-KB locus that includes two genes (*B4GALNT2* and *IGFBP1*) and one pseudogene (*EZR*) and potentially an additional ERV, which were identified in this thesis (see 4.6).

| Gene | Mutation | Allele | Founder breed | Reference |
|---|---|---|---|---|
| *BMP15* | V299D | $FecX^{I}$ | Romney, Inverdale | Galloway et al. (2000) |
| | Q291Ter | $FecX^{H}$ | Romney | Galloway et al. (2000) |
| | S367I | $FecX^{B}$ | Belclare | Hanrahan et al. (2004) |
| | Q239R | $FecX^{G}$ | Belclare, Cambridge | Hanrahan et al. (2004) |
| | Note1* | FecXR | Rasa Aragonesa | Martinez-Royo et al. (2009) |
| | C321Y | $FecX^{L}$ | Lacaune | Bodin et al. (2007) |
| | T317I | $FecX^{Gr}$ | Grivette | Demars et al. (2013) |
| | Note2* | $FecX^{Bar}$ | Barbarine | Lassoued et al. (2017) |
| *BMPR1B* | Q249R | FecB | Boorola Merino | Mulsant et al. (2001) |
| | | | Garole, Javanese | Davis et al. (2002) |
| | | | Small Tailed Han, Hu | Davis et al. (2006) |
| | | | Kendrapada | Kumar et al. (2008) |
| *GDF9* | S39F | $FecG^{H}$ | Belclare, Cambridge | Hanrahan et al. (2004) |
| | S109R | $FecT^{T}$ | Icelandic | Nicol et al. (2009) |
| | F345C | $FecG^{E}$ | Santa lnês | Silva et al. (2011) |
| | V371M | $FecG^{F}$ | Finnsheep | Mullen and Hanrahan (2014) |
| | | | Norwegian White | Våge et al. (2013) |
| | | | Belclare | Mullen et al. (2013) |
| | R315C | $FecG^{V}$ | Note3* | Souza et al. (2014) |
| | R87H | $FecG^{I}$ | Baluchi | Moradband et al. (2011) |
| *B4GALNT2* | Note4* | $FecL^{L}$ | Lacaune | Drouilhet et al. (2013) |
| | | $FecL^{L}$ | D'man | Ben Jemaa et al. (2019) |

### 1.3.2 Nutritional influence on sheep prolificacy

Several factors such as seasonality, environment, day length, and nutrition affect the reproductive performance of sheep (Bermejo et al., 2010; Notter, 2000; Petrovic et al., 2012). Flushing diets (diets with increased nutrient levels) reportedly increase estrous activity and ovulation rate in sheep (Sormunen-Cristian and Jauhiainen, 2002). Moreover, poor nutrition has several negative physiological effects such as weight loss, delayed puberty onset, and infertility (Hernandez-Medrano et al., 2012; Robinson, 1996; Sormunen-Cristian and Jauhiainen, 2002; Ying et al., 2013). However, the genetic and physiologic mechanisms underlying these nutritional influences remain unclear. An in-depth understanding of how a flushing diet influences folliculogenesis and ovulation rates is essential for facilitating targeted nutrition and improving overall fertility in sheep (Scaramuzzi et al., 2011). Thin ewes respond more strongly than those in above-average conditions, and the greatest response is observed early and late in the breeding season. Flushing diets are generally administered as silage and 200–400 g of grain per ewe daily; this increases energy and protein levels in feed. The flushing diet is typically started approximately 2–3 weeks prior to breeding, and it is important to continue administering the same feed for 1 month after breeding to minimize embryonic loss (Chaturvedi et al., 2006). The extra feed ensures that the fertilized eggs have a much better chance of establishing a pregnancy.

## 1.4 RNA-seq as a tool for functional fertility genomics

### 1.4.1 Transcription

The genomic information of living organisms, encoded in their DNA, is expressed through transcription. Briefly, RNA is synthesized from one copy of double-stranded DNA. After transcription, the introns are excised, and exons are ligated to form mature mRNA via RNA splicing. Similarly, Drosha, a key component of the microprocessor complex, cleaves intronic primary miRNA transcripts (usually) from the same pre-mRNA to produce a short molecule ($\sim$ 75-bp) with a stem loop, known as precursor miRNA (pre-miRNA). Further processing of the pre-miRNA yields ($\sim$ 22-bp) double-stranded RNA, one strand of which is incorporated into the RNA-induced silencing complex (RISC) and binds the 3´-UTR of the target mRNA to control protein production either by silencing gene expression via mRNA cleavage or by mediating translational repression (Lee et al., 1993; Bartel, 2004; Kim and Kim, 2007) (Fig. 1.3). The functions of all cells are controlled by the quantitative expression of these protein-coding and non-coding RNA transcripts determined by external and internal factors. RNA transcripts with protein-coding potential are known as mRNAs, whereas the other transcripts that lack potential

to be coded are known as non-coding RNAs (ncRNAs). The collective representation of mRNAs and ncRNAs in a cell at any given time, together with their quantity, is referred to as the transcriptome. This thesis focuses on two classes of RNAs: mRNA and miRNA.

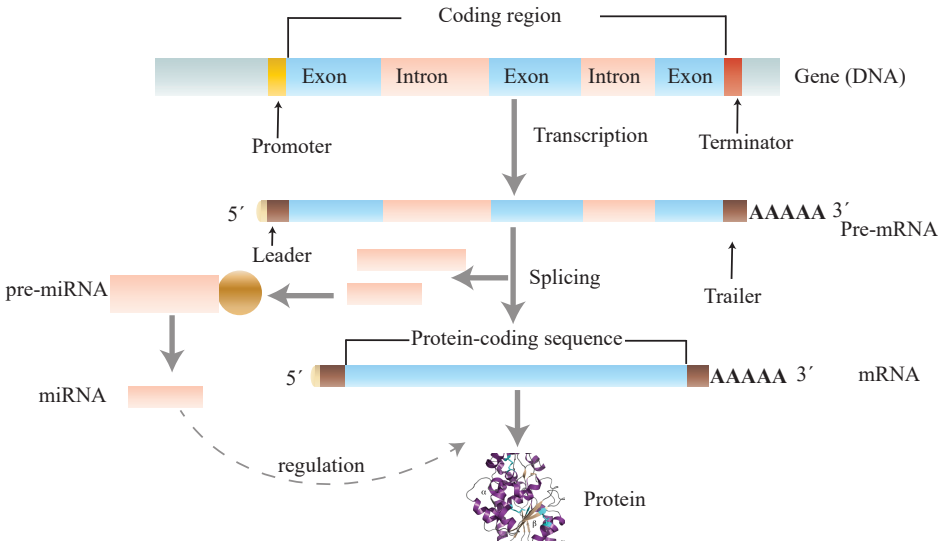


Figure 1.3: Overview of transcription and gene regulation

### 1.4.2 RNA-seq

RNA-seq is a genomic technology that has revolutionized our understanding of the functional aspects of genomes for more than a decade now (Stark et al., 2019; Wang et al., 2009). There are several platforms for RNA-seq; however, Illumina sequencing is the most widely used because of its accuracy (>99%), speed, and high throughput. All sequencing experiments in this thesis were performed using the Illumina sequencing platform based on a sequencing-by-synthesis approach. The basic steps include RNA extraction, cDNA synthesis by reverse transcription, (predefined) adapter ligation, and massively parallel sequencing; a large number of user-defined reads are thus generated for each cDNA library. The sequenced reads are then computationally processed to verify their biological validity. Apart from mRNA and miRNA, sequencing data may also include other classes of RNA, such as ribosomal RNA (rRNA), transfer RNA (tRNA), and other small RNAs. Several tools and pipelines are available for analyzing mRNA (mRNA-seq) and miRNA (miRNA-seq) sequencing data. An overview of the computational aspects of mRNA- and miRNA-seq data are summarized in 1.4.3 and 1.4.4, respectively. For convenience, the two major classes of RNA studied in this thesis are mRNA

and miRNA, and therefore, their sequencing data have been referred to as mRNA-seq and miRNA-seq, respectively.

### 1.4.3 mRNA-seq data analysis

Depending on the research questions and experimental design, various analyses can be performed using mRNA-seq data. Typically, RNA-seq data analysis involves mapping clean reads to an existing reference genome or transcriptome (if one is not available, assembling *de novo*), transcript quantification, and differential expression modelling.

Base-called raw sequence data are typically available as a FASTQ format file and include the sequence ID, barcode (more recently, barcodes can be removed using Illumina's bcl2fastq tool), the sequence itself, and a quality score for each base (Cock et al., 2010). Before further processing of the data, it is important to check the overall quality of the data and the presence of PCR artifacts and other sequences. FASTQC (Andrews, 2019) is one of the most common tools used for this purpose. After checking the quality of the raw sequence data, low-quality data and adapters can be removed using tools such as Trimgalore (Krueger, 2019).

High-quality clean reads are aligned to the annotated genomes of interest to allocate each of the reads to genomic coordinates. Splice-aware mapping tools, such as STAR (Dobin et al., 2013) or HISAT (Kim et al., 2015) (which succeeded TopHat (Kim et al., 2013)), are typically used. In addition, the aligned reads can be assembled into transcripts using assemblers such as Cufflinks (Trapnell et al., 2010) and StringTie (Pertea et al., 2015). Cases in which a reference transcriptome does not exist, the reads can be assembled *de novo*, for example, by using Trinity (Grabherr et al., 2011). Recently, alignment−free tools, such as Salmon (Patro et al., 2017) and Kallisto (Bray et al., 2016), have been developed to directly allocate each sequence read to its respective transcript of origin. Also referred to as pseudo aligners, these tools have been found to be superior (in terms of accuracy) or at least similar to mapping-based tools, with an added advantage of computational efficiency (Everaert et al., 2017; Sahraeian et al., 2017). To mitigate the aberrant alignment between reads from an unannotated genome locus with sequence similarity to the transcriptome, and thus improve the quantification sensitivity, new methods such as selective alignment in Salmon (Srivastava et al., 2020) have been implemented in recent versions of these tools.

To obtain a quantitative overview of the expressed transcripts, all mapped or pseudo-aligned reads must be assigned to their respective transcripts in the reference genome. Some popular tools used for this are HTSeq-count (Anders et al., 2015), RSEM (Li and Dewey, 2011), Salmon, and Kallisto. Typically, expression matrices are created in this step, which can later be utilized for differential expression (DE) analyses.

DE analysis forms the core of RNA-seq-based studies in which the quantified

reads that overlap transcripts are filtered, normalized, and statistically modeled to predict significant changes in gene expression because of functional differences between one or more groups. This step typically requires two tables, wherein one is the matrix of genes in which counts (actual read counts or normalized abundances) for each gene in the first column are recorded for each sample in the first row and the second is a table that includes sample information. DESeq2 (Love et al., 2014) and EdgeR (Robinson et al., 2010) are two Bioconductor packages commonly used in DE analyses. Both tools use a generalized linear model to estimate the overall expression strength of the gene and its ratio ($\log_2$-fold change) between two conditions (e.g., treatment *vs.* control).

Gene annotation assigns biologically meaningful values to expressed gene transcripts. The genes can be annotated individually or by linking them to known Gene Ontology (GO) terms and/or biological pathways. Several genes may typically lack annotation; in such cases, a homology-based approach can be used to determine the annotation from closely related species. The assumption is that genes showing sequence similarity have similar functions. The biomaRt Bioconductor package can be used to retrieve additional information (such as gene name, description, and chromosome location) for one or more genes of interest. Additionally, web-based tools such as DAVID (Dennis et al., 2003) and AgriGO (Du et al., 2010), as well as stand-alone tools such as Cytoscape (Shannon et al., 2003) and Blast2GO (Conesa et al., 2005), are useful for identifying GO terms or KEGG pathways associated with a subset of genes.

### 1.4.4 miRNA-seq data analysis

The analysis of miRNA-seq data slightly differs from that of mRNA-seq data and typically involves five steps: (1) data preprocessing, (2) read mapping, (3) miRNA prediction and quantification, (4) DE analysis, and (5) functional analysis. Numerous analysis pipelines such as CAP-miRSeq (Sun et al., 2014), mirPRo (Shi et al., 2015), and miARma-Seq (Andrés-León et al., 2016) are available for streamlining some or all steps of miRNA-seq data analysis.

Data preprocessing is one of the most important steps in miRNA-seq data analysis. Although the same tools used for mRNA-seq data preprocessing can be used, the parameters for adapter trimming must be adjusted. Some tools also provide automatic settings by specifying that the data are small RNA-seq data as one of the parameters. In any case, the idea is to remove 3´ adapters and low-quality reads, as well as exclude all reads shorter than 17 bases. Alignment tools such as BWA (Li and Durbin, 2009), Bowtie (Langmead et al., 2009), and STAR can map clean miRNA reads to the genome of interest. For computational efficiency, reads with multiple copies may be collapsed before the mapping step so that only one copy will be used during alignment. The mapped reads will be uncollapsed after mapping and used for the quantification and prediction of

known and novel miRNAs. MirDeep2 (Friedländer et al., 2012) is a preferred suite that includes a range of scripts for miRNA quantification and discovery. miRBase (Griffiths-Jones et al., 2008) stores the most comprehensive list of known miRNAs. Therefore, sequences from this database are used to identify the following: (1) species-specific miRNAs (miRNAs that map to already existing miRNAs for a given species), (2) conserved miRNAs [miRNAs that do not belong to (1) but are known in other species], and (3) novel miRNAs (which are predicted to be miRNAs but do not exist in miRBase). Using these steps, we can create a catalog of miRNAs expressed in a given tissue. Next, if the objective is to compare the expression level between more than one group or condition, DE tools such as DESeq2 or EdgeR can be used. Similarly, the regulatory role of miRNAs on target genes can be estimated using tools such as TargetScan (Agarwal et al., 2015).

### 1.4.5 Applications of RNA-seq in sheep genomics

RNA-seq is currently the most popular tool for whole-transcriptome-based studies. While gene expression is routinely studied, numerous other aspects of RNA biology can be studied using the same data. These include alternative splicing (Kornblihtt et al., 2013), ncRNA -mediated gene regulation (Morris and Mattick, 2014), variant detection, and allele-specific expression (Chamberlain et al., 2015; Salavati et al., 2019). RNA-seq has been continuously applied in sheep, with most studies focusing on fertility traits. Table 1.2 summarizes previous studies conducted using RNA-seq in sheep. The main objective of most of these studies was to compare the quantitative gene expression between two or more groups.

Table 1.2: Previous studies performed using RNA-seq in sheep

| Experimental design | Analysis methods | Results | Reference |
|---|---|---|---|
| **Reproduction and fertility** | | | |
| miRNA-seq of follicles from pooled libraries (7–8 ewes each) representing D1, D3, and D9 (D = day) after estrus | quantification and expression, qRT-PCR, northern blotting | 111 miRNAs identified, some miRNAs related to follicular-luteal transition | McBride et al. (2012) |
| Ovarian RNA-seq of high- (Qira black, n = 4) and low-prolific (Hetian, n = 5) breeds | gene expression, GO and KEGG, qRT-PCR | 1252 differentially expressed genes (DEGs), some DEGs associated with steroid hormone biosynthesis | Chen et al. (2015) |
| mRNA and long non-coding RNA (lncRNA) profiling between polytocous (Po, n = 4) and monotocous groups (Mo, n = 4) of Hu sheep | mRNA and lncRNA expression, network (mRNA–mRNA and mRNA–lncRNA), GO and KEGG | Higher levels of LH in the Po group, 76 DEGs and 5 DE lncRNAs, upregulation of lysosomal protein in the Mo group | Feng et al. (2018) |

Table 1.2: Previous studies performed using RNA-seq in sheep

| Experimental design | Analysis methods | Results | Reference |
|---|---|---|---|
| RNA-seq of the testis of three Hu sheep each at D5, 3M, 6M, 9M, and 2Y (D = day, M = month, Y = year) | mRNA and lncRNA expression, network (mRNA–lncRNA), GO and KEGG, and qRT-PCR | lncRNAs, and genes (eg. *CYP19A1*, *SERPINA5*, *PRKCD*) associated with testis development and spermatogenesis | Yang et al. (2018) |
| Gene expression profiles of granulosa cells and oocytes from the ovaries of 12 newborn lambs | laser capture microdissection, microarray, network analysis, qRT-PCR | cell-specific genes, signaling pathways because of cross-talk between two cell types | Bonnet et al. (2011) |
| Ovarian RNA-seq of three ewes each from high- (Small-tailed Han) and low-fecundity (Surabaya fur) breeds while on estrus | gene expression, GO enrichment, and qRT-PCR | *TGFBP1* as high prolificacy candidates, protein synthesis activity related to prolificacy | Miao and Luo (2013) |
| Twelve hypothalamic samples from Po and Mo Small-tailed Han sheep during the follicular (3P and 3M) and luteal (3P and 3M) phase | lncRNA expression, lncRNA-target identification, and network analysis, qRT-PCR | protein synthesis in hypothalamus linked to ovulation rate | Zhang et al. (2019) |
| Ovarian RNA-seq from multiparous (n = 5) and uniparous (n = 5) groups of Mexican Pelibuey breed | gene expression, GO and KEGG, and qRT-PCR | 354 DEGs | Hernández-Montiel et al. (2019) |
| Twelve uterine samples from Po and Mo small-tail Han sheep during the follicular (3P and 3M) and luteal (3P and 3M) phase | mRNA and lncRNA expression, mRNA–lncRNA network. qRT-PCR | 25,104 lncRNAs, 20,908 mRNAs, 16,016 novel transcripts; DEGs associated with ovarian steroidogenesis, retinol metabolism, oxytocin signaling pathway | La et al. (2019) |
| **Tail phenotypes** | | | |
| Three Chinese sheep breeds with different tail phenotypes (long fat-tailed, thin-tailed, and short thin-tailed), three animals in each group | gene and lncRNA expression, lncRNA-gene co-expression, GO, and KEGG | DEGs and lncRNAs associated with enlargement of adipose tissues | Ma et al. (2018) |
| Adipose tissues of three male lambs each from Iranian fat- (Lori-Bakhtiari) and thin-tailed (Zel) breeds | gene expression, GO and KEGG, protein-protein interaction, and qRT-PCR | 264 DEGs (mostly associated with lipid metabolism) | Bakhtiarizadeh et al. (2019) |
| **Muscle development and meat quality** | | | |
| Analysis of meat quality using nine muscle types on four Poll Dorset cross-bred male lambs | *De novo* assembly, DE, GO, and qRT-PCR | DEGs were associated with several properties of meat (e.g., yellowness, tenderness, pH, and water-holding capacity) | Armstrong et al. (2018) |

Table 1.2: Previous studies performed using RNA-seq in sheep

| Experimental design | Analysis methods | Results | Reference |
|---|---|---|---|
| RNA-seq of the longissimus dorsi muscle from three adult ewes and three embryos | lncRNA expression, target prediction, enrichment, and qRT-PCR | varying lncRNA expression at two different stages | Li et al. (2019) |
| | **Milk** | | |
| Milk samples from four Churra and four Assaf ewes, days (D10, D50, D120, and D150) after lambing | variant identification and annotation | variants in genes related to protein and fat content in milk | Suárez-Vega et al. (2017) |
| | **Immunity** | | |
| Three male and three female crosses of Texel and Scottish Blackface (BFxT); one cell type (bone marrow-derived macrophages), six tissues | allele-specific expression (ASE) analysis | ASE analysis pipeline, ASE profiles were tissue- and/or cell-type-specific | Salavati et al. (2019) |
| | **Disease** | | |
| RNA-seq of the udder and spleen from three *M. agalactiae*-infected and three healthy ewes (Austrian mountain breed) | gene expression, qRT-PCR | upregulated DEGs mainly associated with immuno-inflammation | Chopra-Dewasthaly et al. (2017) |
| RNA-seq of lung samples (pooled) from four jaagsiekte sheep retrovirus (JSRV)-infected and four mock-infected sheep | gene expression, comparison with human data, RT-PCR, immunohistochemistry | DEGs associated with immunomodulation and carcinogenesis | Karagianni et al. (2019) |
| Merino ewes (n = 63) grouped by osteotomy models for standard and delayed bone healing, RNA-seq of pooled samples of 4–5 ewes from each group (D7, D11, D14, and D21) after surgery | gene expression, GO, clustering, and qRT-PCR | genes (eg. cytokines) and GO terms (e.g., striated muscle contraction and contractile fiber) associated with bone healing | Jäger et al. (2011) |

# Chapter 2

# Aims of the Study

Farm animal genetic resources have a high socio-economic importance. Studies are warranted to characterize animal genetic resources in order to enable the appropriate and sustainable utilization of native breeds for guaranteeing future food security (Kantanen et al., 2015; Tixier-Boichard et al., 2015). The primary focus of the current work was to characterize the productive traits of Finnsheep, a globally important domestic sheep native to Finland; this thesis also addresses an action plan of the Food and Agriculture Organization (FAO) of the United Nations (UN) for animal genetic resources[1] and is also part of Finland's National Genetic Resources Programme for Agriculture, Forestry and Fishery[2] coordinated by the Natural Resources Institute Finland (Luke).

The overarching aim of this thesis was to apply genomic approaches to gain an insight into sheep reproduction in general and specifically analyze the interplay between relevant reproductive tissues and gene expression within and between high- and low-prolific sheep breeds. The specific aims for this thesis were obtained from four different experimental settings (see 3.2 for details)–pilot experiment, first phase of the main experiment, second phase of the main experiment, and the additional experiment.

**Pilot experiment:** The pilot experiment (I, II) was conducted to evaluate the overall experimental and computational aspects of the main experiment.

**First phase of the main experiment:** The aims of the first phase of the main experiment (III) were to: (a) assess gene expression dynamics in sheep ovaries during the follicular growth phase of the estrous cycle, (b) compare gene expression differences between high-prolific (Finnsheep) and low-prolific (Texel) sheep and

[1] http://www.fao.org/3/a1404e/a1404e00.htm
[2] http://urn.fi/URN:ISBN:978-952-366-182-0

their F1 crosses, (c) assess the impact of a flushing diet on folliculogenesis, and (d) assess the role of miRNA in gene regulation.

**Second phase of the main experiment:** For the second phase of the main experiment, we initially aimed to obtain an insight into the interplay between the endometrium and corpus luteum (CL) during the preimplantation stage of pregnancy. However, after realizing the sampling bias in the endometrium data, the transcriptome profiles of the two tissues were separately analyzed (IV, V). Moreover, miRNAs expressed in the CL and endometrium and their possible roles in gene regulation were assessed (VI).

**Additional experiment:** The main aim of the additional experiment (VII) was to compare gene expression differences between domestic (Finnsheep) and feral (European mouflon) sheep in two reproductive tissues—the ovary and endometrium. The role of miRNA-mediated post-transcriptional gene regulation was assessed using integrated mRNA-miRNA analysis.

# Chapter 3

# Materials and Methods

## 3.1 Breeds under study

Two domestic sheep breeds (i.e., Finnsheep and Texel), their F1 crosses, and feral European mouflon sheep were evaluated. Finnsheep (suomenlammas in Finnish) is native to and is the national sheep breed of Finland (Fig. 3.1). Finnsheep have been reported to be existing since the Iron Age and are survivors of the first migratory event from the Fertile Crescent to Europe (Chessa et al., 2009; Ryder, 1981). Finnsheep are among the most prolific sheep breeds worldwide (Baird and Campbell, 1998) and are exported to more than 40 countries for developing composite/synthetic breeds through cross-breeding programs. Of note, Dolly, the most famous sheep, was cloned using 50% of genetic material from Finnsheep[1]. While most sheep normally breed in autumn, Finnsheep ewes can become pregnant during out-of-season breeding periods as well. Other characteristics of Finnsheep include early-age pregnancy and low lamb mortality[2]. The average litter size of Finnsheep is 2.8 (Fahmy and Dufour, 1988; Maijala, 1988); however, there have been several reports of more than five lambs per litter, including an unofficial world record of nine lambs in a single parturition. Overall, the Finnsheep genome may provide clues to several interesting phenotypes, such as adaptation to cold climates, domestication history, and exceptional prolificacy including year-round lambing.

[1] https://dolly.roslin.ed.ac.uk/facts/the-life-of-dolly/index.html
[2] https://www.finnsheep.org/finnsheep-in-print/finnsheep-in-spin-off/

Figure 3.1: Finnsheep ewe. Photo by Sarita Mikkonen


Figure 3.2: Texel ewe that won the "Best ewe show" award during the OKRA agricultural fair 2018. Photo provided by Paula Kiiski

The Texel breed originated from Texel island, Netherlands during the early $20^{th}$ century from crosses of local breeds with British (e.g., Leicester, Longwool and Lincoln) and Northern European breeds (Arnaud et al., 2007). They are also an important breed of sheep known for their good conformation, muscularity, and carcass leanness (Fig. 3.2). However, owing to their average litter size of 1.5, they are not as prolific as Finnsheep.

Figure 3.3: A male Sardinian mouflon. Photo by Valeria De Gioannis (valeriadegioannis.com)

Feral European mouflons (*Ovis orientalis musimon*) are the only European wild-type sheep species. Although Asiatic mouflons (*Ovis orientalis*) are considered the ancestors of modern sheep (Hiendleder et al., 2002), European mouflons are thought to be the relics of the first sheep migratory event that occurred ∼6–7,000 years ago in the Mediterranean islands of Corsica and Sardinia (Chessa et al., 2009). After readaptation to the feral life, they were introduced across mainland Europe during the 19 $^{th}$ century (Guerrini et al., 2015; Poplin, 1979). Today, mouflons can be found in different regions of Europe, including Finland. The mouflons (current estimate is over 100 animals) in Finland were introduced to two islands, one in the Gulf of Finland and the other on Säppi island located in the Bothnian Bay (Ruusila and Kojola, 2010). Mouflon rams are characterized by spectacular horns, whereas ewes are typically polled (Fig. 3.3).

## 3.2 Experimental design

Animal handling and all procedures of feeding experiments and tissue sampling were conducted under the license approved by the Southern Finland Animal Ex-

periment Committee (approval ESAVI/5027/04.10.03/2012). All experimental domestic sheep were kept at Pusa farm in Urjala (coordinates: 61°05'N 023°33'E), southwest Finland. Biopsy samples from mouflon sheep were collected as soon as possible after being killed by hunters on Säppi island (coordinates: 61°28.6'N 021°21.1'E) on the Bothnian Sea coast. The experiments conducted in this thesis are summarized in Fig. 3.4 and can be categorized as pilot experiment, main experiment, and additional experiment.

The pilot study involved the transcriptome profiling of two Finnsheep ovaries collected during the out-of-season breeding period (I and II).

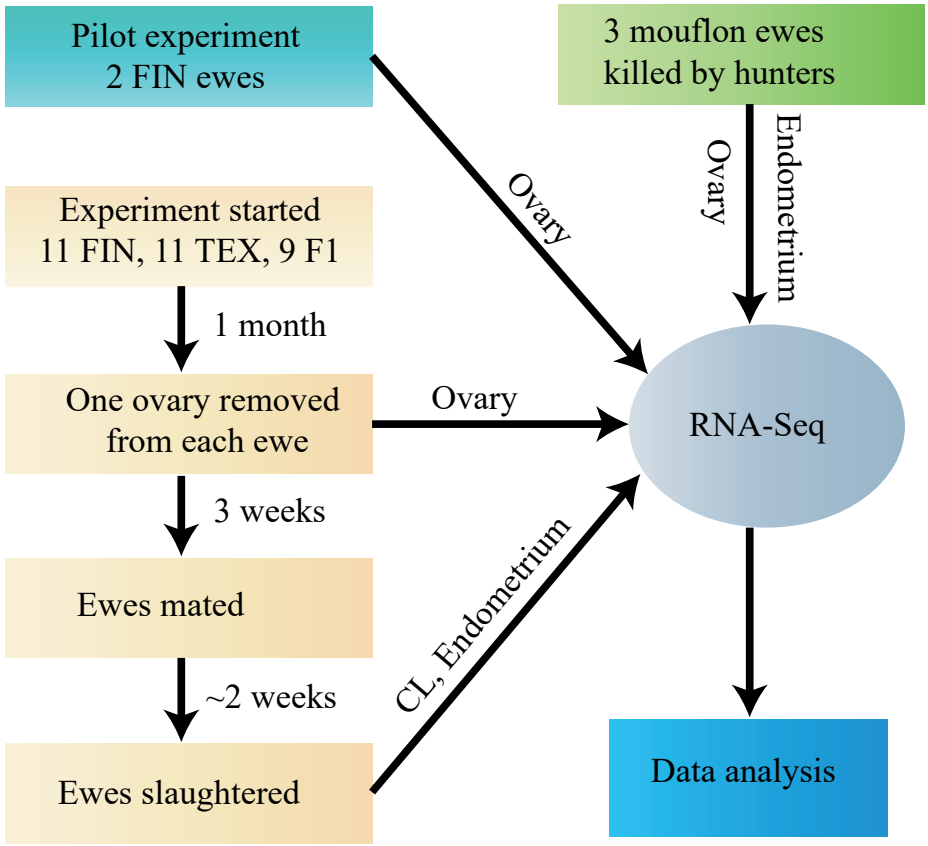


Figure 3.4: **Experimental design**. Abbreviations: FIN - Finnsheep, TEX - Texel, F1 - F1 crosses of FIN and TEX

The main experiment involved RNA-seq of three reproductive tissues: the ovary, CL, and endometrium. To study the nutritional impact, half of the ewes from each breed were maintained on a flushing diet throughout the experiment (for details of feeding experiments, see Additional file 1 of III). The main experiment

was carried out at two different time points (phases) during the establishment of pregnancy: follicular growth phase (first phase) and early pregnancy prior to implantation (second phase). In the first phase, one ovary from each ewe (11 Finnsheep, 11 Texel, and 9 F1) was surgically excised during the follicular growth phase of the estrus cycle. Blood parameters and hormonal measurements (see Additional file 2 in III) were evaluated to follow estrus cycle progression in order to estimate sample collection during the first phase of the experiment. In the second phase, the sheep were mated and slaughtered at different time points (7–17 days after mating) during early pregnancy to collect another set of tissue samples (the liver, embryos, CL, and uterus) from all animals. Among the tissues collected during the second phase, only the CL and uterine epithelia of six animals from each breed were utilized for RNA-seq.

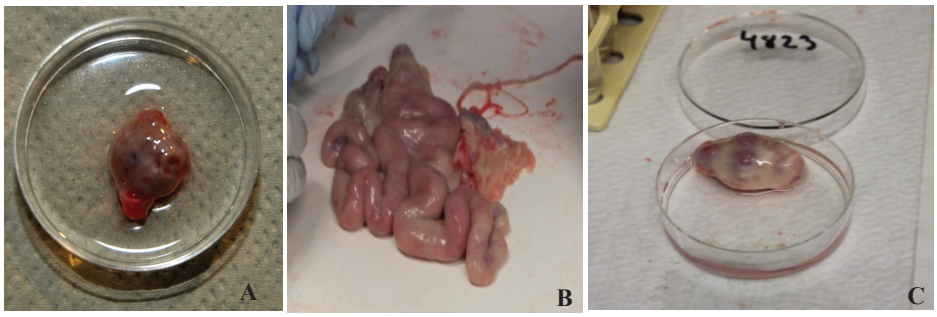


Figure 3.5: **Reproductive tissues under study**. The whole ovary (A) was used in the first phase of the study. The luminal epithelia from the uterine horns (B) and CL from the ovary (C) were used in the second phase of the study

An additional experiment was conducted to understand the evolution of reproduction by comparing the ovarian and endometrial transcriptome profiles of domestic (Finnsheep) and feral (European mouflon) sheep. RNA-seq data of Finnsheep ewes selected from the main experiment were compared with newly sequenced RNA-seq data of mouflon ewes (VII). In this study, ovarian samples from six Finnsheep ewes were compared with the corresponding tissue of three mouflon ewes (six samples, including three technical replicates). Similarly, endometrial RNA-seq data of two Finnsheep were compared with those of one mouflon ewe (two biological replicates representing two uterine horns, see materials and methods section of VII).

## 3.3 Methodological considerations

The development of genomics methods and tools is an active area of research, and it remains difficult to identify a universal tool or standard method for analysing

high-throughput sequencing data. Many studies have employed multiple tools and resources before selecting the most appropriate method. Furthermore, several steps in data analysis are flooded by such choices. This section highlights how such tools, methods, and parameters have been applied to different studies of this thesis. The gene expression profiles in the endometrium tissues were apparently influenced by the physiological changes during embryonic growth, as revealed by the principal component analysis (PCA) plot (Fig. 4.3). Because of this bias, further experiments, such as DE between breeds, were not performed for the endometrial RNA-seq data. Instead we compared the overall gene expression differences in the endometrium samples influenced by spherical and elongated embryos (V). For biopsy, the tissue samples of domestic sheep were collected in a more controlled environment, whereas those of European mouflons were collected after being hunted. In publication II, a 100-bp paired-end strategy was used to sequence miRNAs, and only forward reads were used for downstream analysis. In the remaining experiments, a 50-bp single-end strategy was adopted for miRNA sequencing.

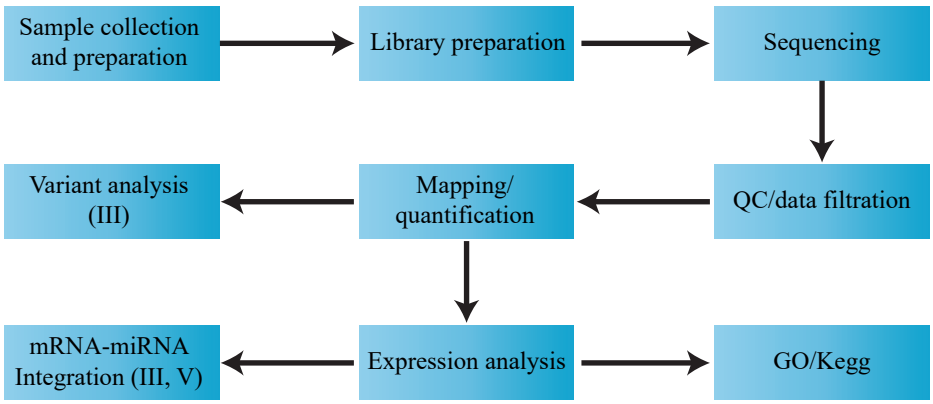


Figure 3.6: **Sequencing and data analysis flowchart**. Minimalistic overview of the sequencing and data analysis

There are some variations in the tools and methods used for RNA-seq data analyses. Because of insufficient sample sizes, DE analyses were not performed in the pilot study (I, II). In III, miRNA-seq data were analysed using CAP-miRSeq, a semi-automated pipeline (Sun et al., 2014). Variant analysis in III did not provide useful data, except that the presence of miRNA cluster in sheep chromosome 18 (C18MC) was first noted in miRNA-seq variant analyses. Variant analyses and qRT-PCR did not reveal any substantial information and were, therefore, not considered in subsequent studies. Majority of the data analyses (I, II, III, V, VI) were performed using the Texel reference genome (Oar3.1). Oar4.0 (the improved version of Oar3.1) and the recent Rambouillet reference genome were used in VII

and IV, respectively. Moreover, the transcriptome of European mouflon sheep was assembled *de novo*. In all cases, gene annotations were based on the Texel reference genome (Oar3.1).

Mapping and quantification tools have received increased attention in RNA-seq data analyses. Tophat2 was mainly used for mapping mRNA reads (I, III and VII); however, in IV (in addition to STAR) and V, the reads were quantified at the transcript level using Salmon. The Salmon counts were later summarised at the gene level using the tximport R package (Soneson et al., 2016), prior to DE analysis. All miRNA-seq reads were aligned using Bowtie. Similarly, another important procedure in RNA-seq data analysis, i.e., DE analysis, was mainly performed using DESeq2, except in VII wherein EdgeR was used to compare gene expression in mouflon tissues (mRNA-seq; ovary *vs.* endometrium), Cuffdiff was used to compare gene expression between species, and DESeq was used for miRNA DE analysis. Significantly differentially expressed genes from the comparisons were selected based on the adjusted p-values (padj) and $\log_2$(fold-change) (LFC) ratios.

Blast2GO was initially used for gene annotation and enrichment analyses (I). After the free version of Blast2GO was discontinued, alternative programs, namely ClueGO (III, IV, V), DAVID (VII), and PANTHER (VII), were used for GO and KEGG pathway analyses. All gene annotations (gene name, gene description, chromosome number, and in some cases, human orthologs) were retrieved using the Bioconductor biomaRt package. Moreover, BLAST was used as a last resort for sequence similarity-based annotations (e.g., novel genes and miRNAs).

The following table summarizes the tools and other resources utilized in this thesis.

Table 3.1: List of tools and methods used in this thesis

| **Methods/Tools** | **Publication** |
|---|---|
| **RNA extraction** | |
| RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) | I–VII |
| **Library preparation** | |
| Illumina TruSeq® total RNA | I, II |
| Illumina TruSeq® stranded mRNA | III, IV, V, VII |
| Illumina TruSeq® small RNA | III, VI |
| **Sequencing** | |
| Illumina HiSeq 2000 | I–VII |
| **Data screening and summarization** | |
| FastQC (Andrews, 2019) | I–VII |
| MultiQC (Ewels et al., 2016) | IV, V |
| Plink (Purcell et al., 2007) | III |
| **Data trimming and adapter removal** | |
| Cutadapt (Martin, 2011) | I–III, VII |
| PRINSEQ (Schmieder and Edwards, 2011) | II, VII |
| TrimGalore (Krueger, 2019) | IV, V, VI |

Table 3.1: List of tools and methods (continued)

| | |
|---|---|
| **Alignment and assembly (mRNA data)** | |
| TopHat2 (Kim et al., 2013) | I, III, VII |
| STAR (Dobin et al., 2013) | IV |
| Cufflinks (Trapnell et al., 2010) | I, III, VII |
| Trinity (Grabherr et al., 2011) | VII |
| **Mapping and discovery of miRNAs** | |
| Bowtie (Langmead et al., 2009) | II, III, VI |
| MirDeep2 (Friedländer et al., 2012) | II, III, VI |
| miRBase (Griffiths-Jones et al., 2008) | II, III, VI, VII |
| CAP-miRSeq (Sun et al., 2014) | III |
| **Quantification** | |
| RSEM (Li and Dewey, 2011) | VII |
| HTSeq-count (Anders et al., 2015) | III, IV |
| Salmon (Patro et al., 2017) | IV, V |
| **mRNA-miRNA integration** | |
| TargetScan (Agarwal et al., 2015) | II, III, VII |
| **Variant analysis and annotation** | |
| SAMTOOLS (Li et al., 2009) | III |
| IGV (Thorvaldsdóttir et al., 2013) | I, III |
| GATK (McKenna et al., 2010) | III |
| VEP (McLaren et al., 2016) | III |
| **Gene/transcript annotation** | |
| Blast2GO (Conesa et al., 2005) | I |
| Cytoscape (Shannon et al., 2003) | III, IV, V |
| ClueGO (Bindea et al., 2009) | III, IV, V |
| biomaRt (Durinck et al., 2009) | III, IV, V |
| BLAST (Madden, 2003) | I, III–V, VII |
| PANTHER (Mi et al., 2016) | VII |
| DAVID (Dennis et al., 2003) | VII |
| **Differential expression analysis** | |
| DESeq (Anders and Huber, 2010) | VII |
| DESeq2 (Love et al., 2014) | III–VI |
| EdgeR (Robinson et al., 2010) | VII |
| Cuffdiff (Trapnell et al., 2010) | VII |
| **Validation experiment** | |
| qRT-PCR | II, III |
| Sanger sequencing | III |

# Chapter 4

# Results and Discussion

## 4.1 Sequence data

Altogether, 853 gigabytes of compressed raw data (fastq.gz) were generated for seven studies representing three experimental settings. The mRNA- and miRNA-seq data along with sample information have been uploaded to the European Nucleotide Archive. The primary accession codes associated with the published data are as follows: PRJEB34419 (I), PRJEB34418 (II), PRJEB22101 (III), PRJEB32852 (IV, V, VI), and PRJNA451237 (VII).

## 4.2 Pilot study (I, II)

A total of 14,875 genes and 304 mature miRNAs (including 60 novel sheep miRNAs) were expressed in the two ovary samples included in the pilot study. As the biopsy of the ovary was performed during the out-of-season breeding period, several genes (*NTRK2*, *NID1*, *FOXO1*, *HTRA1*, *SERPINE2*, and *PPAP2B*) potentially associated with out-of-season breeding (Chen et al., 2012) were expressed. A sequence similarity search revealed that the sequences of these genes were mainly similar to sequences from other mammals (cow, human, dolphin, killer whale, pig, yak, giant panda, wolf, white-cheeked gibbon, bat, etc.), pathogens (e.g., Orf virus), and even insects (silkworm). These findings corroborated those of other studies (Sangiovanni et al., 2019; Whitacre et al., 2015), providing important information for improving the existing annotations and annotating novel genes. Therefore, it is very important to consider distantly related species during the genome annotation process. In particular, the presence of Orf virus (the third most common species in the BLAST result, see Fig. 3 in I) sequences in the data is of great interest. It is likely that the genome of the Orf virus existed either as a retroelement or as a pathogen (Sangiovanni et al., 2019; Usman et al., 2017).

## 4.3 Phenotype observations (III,IV,VII)

All 31 domestic ewes included in the main experiment were mature, with an average weight and age of 71.2 kg and 4.25 years, respectively. Pedigree records and distance matrix based on identity by state (IBS ) from 700k SNP markers of the 31 ewes showed that the F1 crosses had a 50% genetic composition of Finnsheep and Texel each (III, also see Fig. S1 therein). Of the three European mouflons, one was estimated to be 5–6 years old, whereas the other two were younger (∼1.5 years). The reproductive stage of the mouflon ewes was not known, except for the fact that the older mouflon had a CL. One of the key findings of the diet experiment was that the body condition score (BCS) of the ewes on the flushing diet did not change significantly. Over 6 weeks, the average increase in the BCS was 0.26, which is much lower than the typically expected BCS increase of 1 during this period (Dillon et al., 2002).

The number of CLs and embryos appeared to be closer to the possible litter-sizes as compared to the ovarian follicles (Table 4.1). Nevertheless, on average, Finnsheep showed the largest number of follicles. Interestingly, the embryo survival rate (the difference between the number of CLs and embryos) was higher in Texel sheep than in Finnsheep. On average, 2.6 embryos were obtained from 4.1 CLs in Finnsheep (63%), whereas 1.5 embryos were obtained from 1.7 CLs in Texel sheep (88%). Similarly, the 1.8 embryos obtained from 3.75 CLs in F1 indicated the lowest (48%) embryo survival rate among the three breeds. The general observation that ewes with a higher ovulation rate also showed a greater embryo mortality rate corroborates the findings of previous reports (Rhind et al., 1980; Silva et al., 2016). To validate the accuracy of these results, more samples are required.

Table 4.1: **Key phenotype records**. The average number of ovarian follicles, CLs, and embryos in three breeds of domestic sheep. The number of CLs and embryos in pure breeds were mostly proportional to the average litter size estimates, whereas the number of follicles was rather random. For F1 crosses, the number of CLs was closer to that for Finnsheep, whereas the number of embryos was more similar to that for Texel ewes

| | Finnsheep | | Texel | | F1 | |
|---|---|---|---|---|---|---|
| | Flushing | Control | Flushing | Control | Flushing | Control |
| Follicles | 11 | 12.8 | 9.2 | 10.2 | 9.5 | 7.5 |
| CLs | 4.1 | | 1.7 | | 3.75 | |
| Embryos | 2.6 | | 1.5 | | 1.8 | |

## 4.4 Highly expressed genes and their significance (III, IV, V)

Highly expressed genes reflected the shared and unique activities of the three reproductive tissues examined in this thesis. A total of 16,402, 21,287, and 21,125 genes were expressed in the ovary, CL, and endometrium, respectively. The top 25 expressed genes in each tissue were dominated by mitochondrial genes (12 each in the ovary and CL, 9 in the endometrium) in all three tissues. The abundance of the mitochondrial genes was proportional to the energy requirement of the given tissues. In humans, the heart and brain are known to have a high proportion (∼30%) of mitochondrial genes, whereas tissues involved in reproduction (e.g., the ovary, testes, and prostate) are composed of only ∼5% of mitochondrial transcripts (Ali et al., 2019; Mercer et al., 2011). The high abundance of mitochondrial genes in the reproductive tissues (e.g., 14% in the ovary) indicated that female reproduction-related tissues in sheep (polyovulatory species) could be considered as high-energy tissues. Of the top 25 highly abundant genes, nine were shared by all three tissues, reflecting that numerous genes are abundantly expressed in many tissues (Ramsköld et al., 2009). Similarly, 15 genes were shared by the CL and endometrium, indicating their common role during preimplantation.

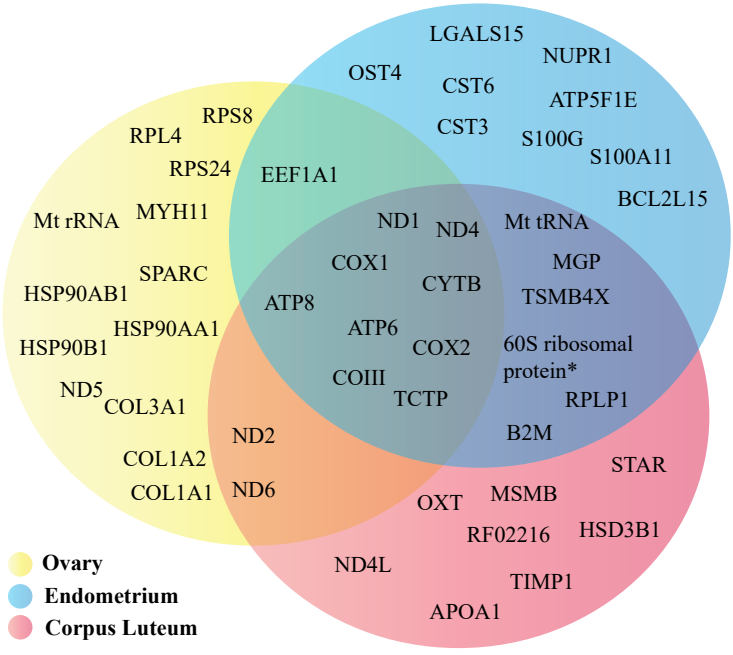


Figure 4.1: Venn diagram of the top 25 abundant genes in the ovary, endometrium, and CL (III, IV, V)

Highly expressed autosomal genes specific to folliculogenesis were associated with proliferation (*SPARC*), elongation (*EEF1A1*), and functions attributed to ribosomal proteins (*RPL4*, *RPS8*, and *RPS24*). Similarly, genes highly expressed in the CL and endometrium were associated with implantation-related activities (conceptus elongation, attachment, migration, etc.), a finding that corroborates those of earlier studies (Brooks et al., 2016; Song et al., 2006; Spencer et al., 2016). Genes (*STAR* and *HSD3B1*) involved in P4 synthesis (Davis and LaVoie, 2018) and P4-induced genes (e.g., *LGALS15*, *CST3*, and *CST6*) (Spencer et al., 2016) were among the most abundant genes. Moreover, based on the findings of Liu et al. (2012), the high abundance of *S100A1* in the data indicated successful reproduction.

## 4.5 Gene expression during the follicular growth phase (III)

### 4.5.1 Relatedness of F1 ewes to pure breeds

The gene expression profiles of F1 crosses in terms of DEGs revealed their close relatedness with Finnsheep. Phenotypic records from the second phase (IV, V) of main the experiment also indicated the same (see Table 4.1). Comparisons involving all three settings (between the flushing group, between the control group, and between groups with diet as the second factor) revealed fewer DEGs between Finnsheep and F1 compared to that between Texel and F1 (III). For example, a comparison of F1 crosses involving diet as the second factor revealed 68 DEGs with Texel sheep but only five DEGs with Finnsheep. Furthermore, F1 crosses shared a larger proportion of the top expressed genes with Finnsheep (n = 25) compared to that with Texel sheep (n = 15). As mentioned in 4.3, F1 ewes derived 50% of their genome each from Finnsheep and Texel sheep; hence, the gene expression-based observation cannot be explained by Mendelian inheritance. Parent-of-origin effects (POEs), such as imprinting, genetic coadaptation, and other paternal effects, probably underlie the transcriptomic dominance of Finnsheep on F1 crosses (Lawson et al., 2013; Mott et al., 2014; Oreper et al., 2018). From the experimental perspective, additional samples, including further crosses and backcrosses, are useful. Similarly, allele-specific expression analyses of the cross-breeds would reveal genes relevant to the POE (Guillocheau et al., 2019; Salavati et al., 2019).

### 4.5.2 Influence of flushing

The results from differential gene expression analysis revealed that the effects of the flushing diet were most pronounced in Texel ewes followed by F1 crosses; in contrast, a null effect was observed in Finnsheep ewes. A total of 118 genes were differ-

entially expressed between the flushing and control groups of Texel ewes, with most genes (n = 71) being upregulated in the flushing group. Similarly, 25 genes were differentially expressed in F1 crosses, with only four genes upregulated in the flushing group. In contrast, the same comparison in Finnsheep did not reveal any DEGs. Comparative gene expression between breeds by using diet as a second factor revealed 38 DEGs between the pure breeds, with most (n = 34) genes upregulated in Texel ewes due to diet differences. Interestingly, 28 (*ABLIM3*, *ADAMTSL5*, *CEACAM16*, *COBLL1*, *CST6*, *CYP7B1*, *DCBLD2*, *EHD3*, *EPAS1*, *FBLN2*, *FICD*, *GRAMD1B*, *LPAR2*, *MCFD2*, *MFAP5*, *NPY1R*, *PAPPA*, *PLXNB3*, *SCARB1*, *SEMA3F*, *TC2N*, *TSPAN2*, *ENSOARG00000005699*, *ENSOARG00000012224*, *ENSOARG00000014207*, *ENSOARG00000002778*, *ENSOARG00000017163*, and *ENSOARG00000000531*) of the between-breed (Finnsheep *vs.* Texel) DEGs overlapped with those of the within-breed (Texel, flushing *vs.* control) DEGs; this yielded a list of candidate genes for further validation. Similarly, 68 DEGs were identified between Texel and F1 of which 44 were upregulated in Texel. The least (n = 5) number of DEGs were found in Finnsheep and F1, with *CA5A* and *CSRP3* upregulated in F1 crosses.Thus, gene expression comparisons revealed that the flushing diet did not influence the highly prolific Finnsheep; however, low-prolific Texel ewes were affected by additional nutrition, which may have led to an increase in the ovulation rate.

### 4.5.3 *GDF9* mutation

No known major candidate genes linked to prolificacy (see Table 1.1) were differentially or abundantly expressed in any breed-wise comparisons made in this thesis. Polymorphisms in *GDF9* were visually assessed using IGV, which revealed the presence of a mutation (rs403536877, C-T transition at 5:41841285), leading to a change in valine (V) at position 371 to methionine (M). Referred to as $\text{FecG}^{\text{F}}$, this polymorphism has been reported in Finnsheep (Mullen and Hanrahan, 2014), Norwegian White (Våge et al., 2013), and Belclare (Mullen et al., 2013) sheep. However, all of these breeds were developed upon crossing with Finnsheep, suggesting the origin of $\text{FecG}^{\text{F}}$. *GDF9* screening on a larger population of Finnsheep (n = 31) and Texel (n = 32) ewes using Sanger capillary PCR revealed an $\text{FecG}^{\text{F}}$ mutation frequency of 0.70 in Finnsheep, corroborating the RNA seq data.

In addition to *GDF9* mutation, we observed a SNP (rs413226920; D129E) in the haemoglobin subunit beta (*HBB*) gene where three Finnsheep, two F1 crosses, and none of the Texel ewes carried this mutation. *HBB* was also significantly upregulated in Finnsheep compared to Texel, and thus, was considered a candidate gene for increased ovulation rate in Finnsheep. Larger population level screening of these mutations will be useful for evaluating their effects and consequently, validating the observations made in this study.

## 4.6 Gene expression dynamics during preimplantation (IV, V)

Crosstalk between the CL and endometrium was observed, and the characteristic role of each tissue was defined based on the differential gene expression profiles. Fifteen of the top 25 highly expressed genes were common in the CL and endometrium (see 4.4). Additionally, 5,393 genes were differentially expressed (LFC $> \pm 2$) between these two tissues. While a significant difference in gene expression profiles was illustrated by tissue-specific clusters (Fig. 4.3), these results may have been influenced by the within-tissue variation of the endometrium samples (see 4.6.2) and should be considered with caution. Among the DEGs, several genes belonging to the same or similar families of genes or proteins were present (some such groups have been summarized in Table 4.2), indicating their functional prevalence or significance.

Table 4.2: Selected DEGs that belong to functionally similar groups of genes

| DEG groups | Upregulated in | LFC range |
|---|---|---|
| Protocadherins (16) | CL | 2.3–10.5 |
| EHD proteins(3) | CL | 3.3–6.3 |
| Cilia- and flagella-associated proteins (4) | endometrium | 2.1–5.2 |
| Desmosomomes (7) | endometrium | 2.9–9.4 |
| Homeobox A (7) and B (10) | endometrium | 2.2–9.9 |
| Homeobox C (4) and D (1) | CL | 4.7–9.5 |
| TGFBs and their receptors (8) | CL | 2.0–4.9 |
| Thrombospondins (10) | CL | 2.8–8.6 |
| Guanylate binding proteins (7) | CL | 2.0–3.1 |
| Plakophilins (3) | endometrium | 3.0–10.2 |
| ERV group K (3) | endometrium | 8.1–10.4 |

Some other handpicked genes upregulated in the CL included *STAR* (LFC = 10), FSH receptor (*FSHR* and LFC = 10), Inhibin A (*INHA*), *PEG10* (paternally imprinted gene, LFC = 9.8), period circadian regulator 3 (*PER3*), and *BMPR1B* (LFC = 3.6). Moreover, genes involved in progesterone formation were upregulated in the CL, with *STAR* and *HSDB31* ranking among the top 25 most abundantly expressed genes. In general, a group of DEGs from several families were shared between the two tissues. However, the CL appeared to have more tissue-specific DEGs. One reason for this is that the CL tissues are more homogenous than the endometrium tissues.

In addition to the three transcripts belonging to endogenous retrovirus (ERV) group K, which were significantly upregulated in the endometrium (Table 4.2),

several transcripts related to ERVs were expressed and there were comparatively more in the endometrium samples than in the CL samples; however, a majority of the ERV transcripts were classified as novel genes. One notable transcript (*ENSOARG00000009959*), originally classified as a "novel gene," was predicted to be an ERV, which is part of the reduced FecL locus known to affect prolificacy in Lacaune sheep (Drouilhet et al., 2013). A BLAST search using the coding sequence of the novel transcript against the NR database resulted in several hits, including the sequence that is part of the FecL locus (gene bank ID: *KC352617*), the sequence from bighorn sheep chromosome 17 (gene bank ID: *CP011902.1*, *Ovis canadensis*), and the partial sequence of ovine endogenous-virus beta-2 pro/pol region (gene bank ID: *AY193894.1*). More than 95% of pairwise sequence similarity as well as the multiple sequence alignment indicated that the novel transcript is possibly an ERV and is part of the FecL locus (Fig.4.2). Previous studies have reported the role of ERVs during pregnancy, including protecting the uterus against pathogens, involvement in blastocyst development, and contribution to placentation (Dunlap et al., 2006; Spencer et al., 2007; Spencer and Palmarini, 2012). They have also been utilized to trace the domestication events and evolution of different sheep breeds (Chessa et al., 2009). At least 32 ERVs integrated into the sheep genome have been identified (Sistiaga-Poveda and Jugo, 2014); however, there are likely many which remain to be characterized.


Figure 4.2: **Identification and characterization of a novel ERV in the FecL locus**. The novel ERV is flanked by *B4GALNT2* and ezrin-like protein in the reduced FecL locus (A). Multiple sequence alignment based on three of the BLAST hits compared the sequence similarity of the novel ERV with the partial sequence of known ERV (kERV), reverse complement of FecL locus (RFecL), and 43U isolate from Chr17 of bighorn sheep (OC43U) (B). The figure is an expanded version of the one available in V.

As the pregnant ewes were not slaughtered on a specific day of pregnancy, the gene expression profiles in the endometrium samples were severely affected, where-

as those in the CL remained more or less stable. This phenomenon was revealed by the PCA plot of the top 500 genes expressed in the two tissues (Fig. 4.3). Based on the PCA summary and phenotype records (Table 1 in V), two subclusters represent the endometrial transcriptome in response to spherical (day 7 to day 12, lower right) and elongated (day 13 to day 17, upper right) embryos. In particular, the upper cluster formed by the samples obtained after day 12 is influenced by the gene expression changes resulting from the synthesis and secretion of IFNT as well as prostaglandins and cortisol from the trophectoderm (of an elongating conceptus). IFNT, in particular, triggers/induces the expression of several implantation-related genes and IFN-stimulated genes in the endometrium (Brooks and Spencer, 2015).

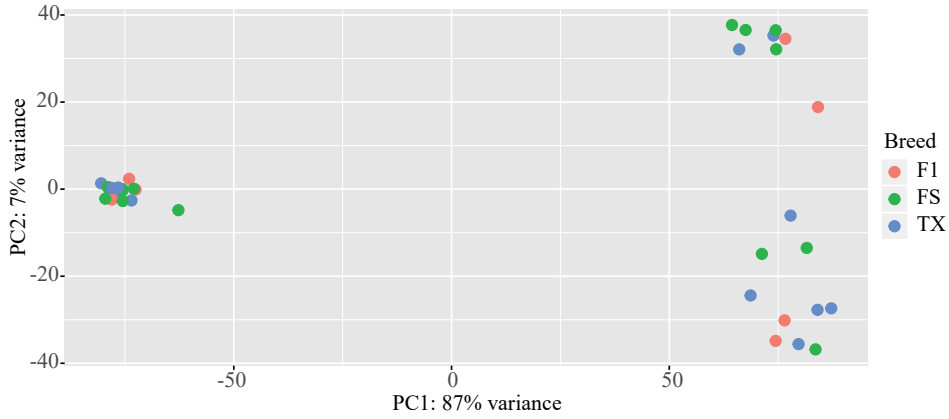


Figure 4.3: PCA based on the top 500 genes expressed in the CL and endometrium. Endometrial samples (right) were grouped into two main clusters representing samples before (7–12 days after mating, upper cluster) and after (13–17 days after mating, lower cluster) the elongation phase of embryo development in sheep. In contrast, the gene expression profiles of the CL samples (left) were relatively homogeneous.

### 4.6.1 Breed specific DEGs in the CL

The highest number of DEGs was observed in the comparisons involving pure breeds (n = 199). Multidrug receptor proteins (n = 8) were upregulated in Texel sheep, whereas a group of sialic acid-binding Ig-like lectins (Siglecs, n = 6) known to play an important role in the immune system (Crocker et al., 2007; Pillai et al., 2012) were upregulated in the CLs of Finnsheep. Most DEGs were upregulated in Finnsheep compared to that in Texel sheep and were mainly associated with the GO terms and KEGG pathways related to implantation (”receptor internalization,” ”cell adhesion,” and ”cell differentiation”) and the immune system (”defense response” and ”immune system process”). One of the interesting genes identified from the DE analyses is *CA5A*, which was always upregulated

in the F1 crosses of samples from both (ovary and CL) the phases of the main experiment. Previous studies have reported the role of carbonic anhydrases in cellular processes in different tissues, including the ovary (Hynninen et al., 2004). In humans, alterations in *CA5A* and its deficiency are reportedly associated with hyperammonemia in children (van Karnebeek et al., 2014). A report that stated its absence in CL (Friedley and Rosen, 1975) has been dismissed by these results. Recently, Hernández-Montiel et al. (2019) also reported that *CA5A* is upregulated in a prolific group of Pelibuey ewes. Thus, *CA5A* plays an important role in reproduction and is of significance in cross-breeding programs; therefore, further characterization of this gene is essential.

### 4.6.2 Endometrium transcriptome in response to embryo development

As shown in Fig. 4.3, endometrial samples were grouped into two sub-clusters, reflecting the gene expression changes prior to (day 7 to day 12) and after (day 13 to day 17) the embryo elongation phase during the course of pregnancy establishment. Anatomically, the embryos representing these two groups of samples were either spherical (S) or elongated (E) and thus, the endometrial samples were referred to as S and E, respectively. We identified 1,410 DEGs between these groups, 554 of which were upregulated in E and included several interferon-stimulated genes, galectins, chemokines, and other immune-related genes (V). The endometrial transcriptome of E was enriched for immune-related processes, whereas tissue morphogenesis- and development-related processes were associated with that of S. Functional aspects of genes upregulated S were affected by limited (e.g., >50% of genes lacked gene descriptions) annotations, which also indicated the knowledge available for previous stages of embryo development.

## 4.7 Generation of a sheep miRNA resource (III, VI, VII)

To our knowledge, the largest amount of sequence data for sheep miRNA (as of April 2019) have been generated. Importantly, the miRNAs of European mouflons were sequenced and quantified for the first time, which included 215 miRNAs in the ovary and 136 miRNAs in the endometrium. In domestic sheep, we identified 304, 342, and 499 sheep miRNAs in the pilot study (II), first (ovary, III), and second phases of the main experiment (CL and endometrium, VI), respectively [1]. The regulatory role of miRNAs in target genes was assessed by integrating differentially expressed miRNAs and mRNAs in two possible groups—Finnsheep *vs.*

[1]Note: For the final counts, miRNAs with reads $\leq 10$ were ignored in II and III, whereas miRNAs $\leq 5$ were ignored in VI

Texel sheep—given the flushing diet (see Fig. 4 in III) right from the first phase of the main experiment (i.e., ovarian RNA seq) and Finnsheep *vs.* European mouflon ovaries from the additional experiment (VII). A comparison of Finnsheep and mouflon ovaries revealed 53 significantly differentially expressed miRNAs, 30 of which were upregulated in Finnsheep and the rest in European mouflon sheep (see VII and Fig. 4 therein). However, the regulatory role of miRNAs relevant to early pregnancy was not clearly understood owing to the low number of differentially expressed miRNAs. Only two miRNAs, both upregulated in Finnsheep as compared to that in Texel sheep, were differentially expressed in the CL (VI), whereas five endometrial miRNAs were differentially expressed between Finnsheep and European mouflon sheep (VII).

The number of miRNAs (>500) quantified in this thesis account for more than three-folds of the number of miRNAs available in the miRBase database (n = 153). Moreover, there is a clear need to enrich the miRNA resource for sheep. miRBase, the primary repository for known miRNAs, contained no additions to the list of sheep miRNAs in its recent update (miRBase version 22). In contrast, the list of miRNAs for species such as cattle (28.8%) and chicken (24.2%) was updated. Given that currently 2,656 mature miRNAs are known in humans and 1,030 miRNAs are known in cattle, the number of miRNAs quantified in this thesis is valuable. Annotation of the non-coding RNAs (including lncRNAs and miRNAs) in different species is currently underway, and the data from this thesis have been utilized by the Functional Annotation in Animal Genomes (FAANG) consortium's non-coding RNA annotation working group. More importantly, when the project kicked off with a hackathon to compile the analysis pipeline and data collection for small RNAs, the data from this thesis contributed to more than 80% of the total sheep miRNA-seq data.

### 4.7.1 Imprinted miRNA cluster on chromosome 18

One of the key findings from the miRNA-seq data analysis was the identification of a large cluster of miRNAs on chromosome 18 (C18MC). In III, all 51 miRNAs known to be present within a 194-Kb region (64,466–64,600 Kb) on chromosome 18 were expressed in the ovarian miRNA-seq data. However, the main cluster was found to be within a 40-Kb region and included 46 miRNAs (Fig. 4.4) and was thus referred to as C18MC. The homologous region of sheep C18MC has already been identified and/or characterized in the genomes of selected placental mammals, including human (C14MC, Bentwich et al. (2005)), mouse (C12MC, Seitz et al. (2004)), and horse (C24MC, Dini et al. (2018)). However, the C18MC homologous regions of several other mammalian species remain to be characterized. Using the Ensembl region comparison tool, the C18MC homologous region was identified in additional species using human C14MC as a query. A similar cluster was identified in the genomic region of elephant (Sc9MC; here, Sc refers to scaffold), dog

(C8MC), goat (C21MC), wild yak (CJH880931.1), pig (ScAEMK02000452.1MC), cow (C21MC), and others. The numbers of miRNAs in the clusters of these species were always lower than the numbers of those in sheep (n = 46), the closest being 44 in mouse C12MC. This high number of miRNAs clustered within an 40-Kb region makes C18MC the largest miRNA cluster described in any species to date. All miRNAs from C18MC were expressed in the first phase of the main experiment (III), whereas three miRNAs (*oar-mir-329a*, *oar-mir-1193*, and *oar-mir-656*) were missing in the second phase tissue samples (VI). In the second phase samples, of the 43 miRNAs expressed, six miRNAs (*oar-mir-411b*, *oar-mir-376d*, *oar-mir-154a*, *oar-mir-496*, *oar-mir-377*, and *oar-mir-412*) were exclusively expressed in the CL. In other words, 43 miRNAs were expressed in the CL, whereas 37 were present in the endometrium.

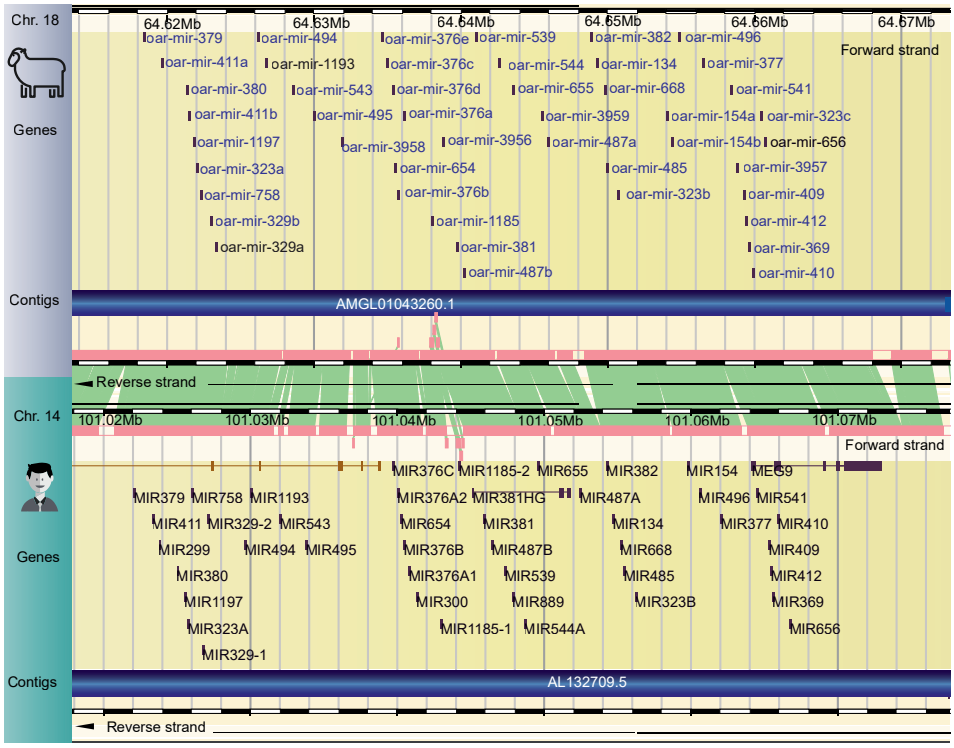


Figure 4.4: **Pairwise genomic alignment of sheep C18MC and human C14MC**. All miRNAs expressed during the preimplantation phase of data collection (CL and endometrium, VI) in sheep (upper half of the figure) are marked in blue text, whereas those not expressed are marked in black text. All miRNAs were expressed in the ovaries from the first phase of the experiment (III). The alignment was performed within the Ensembl browser using Lastz (Harris, 2007), and the original image was modified using Adobe Illustrator. For clarity, some tracks from the human chromosome (e.g., regulatory build) were removed. The gene names for sheep are based on Ensembl release 87, whereas those for human are based on Gencode 29.

To date, three miRNA clusters are known to be imprinted, of which primate-specific C19MC (human) and rodent-specific C2MC (mouse) are known to be paternally imprinted, whereas C14MC (human), a maternally imprinted cluster, is highly conserved among placental (eutherian) mammals (Glazov et al., 2008; Noguer-Dance et al., 2010; da Rocha et al., 2008). These miRNA clusters are known to be regulated by imprinted regions, such as the DLK1/DIO3 domain located ∼200-kb upstream of mouse C12MC and human C14MC (da Rocha et al., 2008; Seitz et al., 2004). Thus, it was concluded that since C18MC in sheep is homologous to C14MC in humans, the former is also a maternally imprinted cluster.

## 4.8 Transcriptome profiles of domestic and feral sheep (VII)

High-quality (N50 length of 700 bp) *de novo* assembled transcriptomes of European mouflon sheep yielded improved results compared to those based on the sheep reference genome. The top expressed genes in the studied tissues (ovary and endometrium) of mouflon ewes were dominated by ribosomal proteins, whereas those from domestic Finnsheep were dominated by mitochondrial genes. There may be at least two reasons for this difference. From the reproduction perspective, two of the three mouflon ewes were young and may have not been under the reproductive cycle. One conclusion derived from this thesis was that in sheep, reproduction is an energy-intensive activity, and this may have led to the abundance of mitochondrial genes (III, V). Another reason underlying the abundance of ribosomal proteins in mouflon ewes may be associated with their adaptation to harsher conditions because of the Finnish winter. Numerous studies have shown that ribosomal proteins play an important role in surviving and adapting to environmental stress (Thieringer et al., 1998; Yurchenko et al., 2018). Recently, several ribosomal proteins were found to be rapidly evolving (expanded) in reindeer (Weldenegodguad et al., 2020) that have adapted to northern Eurasian environments.

Altogether, 192 ovarian and 1065 endometrial mRNA transcripts were significantly differentially expressed between Finnsheep and European mouflon. In contrast, the number of differentially expressed miRNAs were higher in the ovaries (53) than in the endometrium (5). Several genes associated with prolificacy (e.g., *INHBA*, *SERPINA5*, *PTGS2*, *LDLR*, and *EREG*), early maturity (e.g., *CYP19* and *FST*), and out-of-season breeding (*CYP19* and *SERPINE2*) were upregulated in Finnsheep reproductive tissues (ovary and endometrium). However, the genes (*ASTL*, *ZP2*, and *ZP4*) involved in negative regulation of fertilization (Avella et al., 2016; Burkart et al., 2012) were upregulated in European mouflon samples.

Biological processes associated with upregulated endometrial genes in Finnsheep were related to metabolic and biosynthetic processes, whereas those from Euro-

pean mouflon sheep were related to developmental processes. One recent study by Mesquita et al. (2015) on the bovine endometrium during early diestrus showed that the biological processes associated with biosynthetic and metabolic processes were enriched in cows treated for producing larger follicles (≥13 mm), whereas those related to cell proliferation were enriched in normal cows (≤12 mm). Similarly, transcripts associated with metabolic processes and ATPase activity were upregulated in fertile cows compared to those in infertile ones (Minten et al., 2013).

In conclusion, the biological processes enriched in the Finnsheep endometrium were associated with better receptivity and probably higher prolificacy, whereas those enriched in European mouflon sheep indicated a low receptivity phenotype. Integrated analysis based on DEGs and miRNAs revealed causal regulatory networks between miRNAs and their target genes (see Fig. 4 in VII). Important genes, miRNAs, and biological processes relevant to reproduction and fertility are summarized in Fig. 4.5.

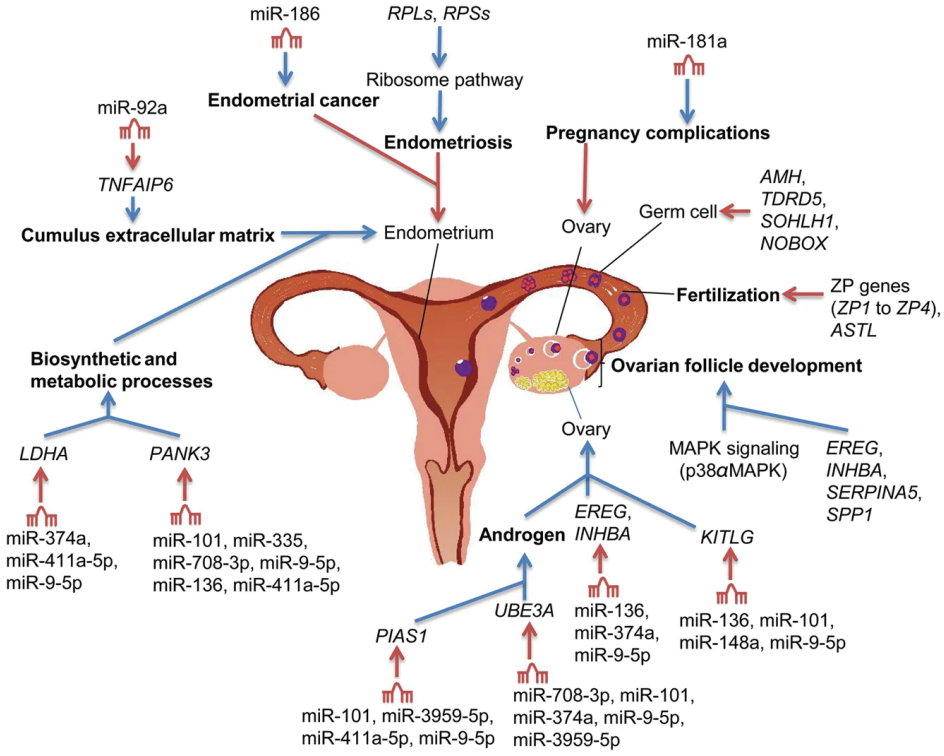


Figure 4.5: **Summary of the key genes and miRNAs relevant to reproduction**. Figure from VII.

# Chapter 5

# Conclusions

Proper genomic characterization is essential for the conservation and sustainable utilization of farm animal genetic resources to guarantee future food security while mitigating challenges such as globalization, climate change, and disease outbreak. This thesis explored the genetic basis of prolificacy, which is of high economic importance in sheep farming systems. High-throughput mRNA and miRNA sequencing of key reproductive tissues captured gene expression dynamics during ovulation (ovary) and early pregnancy prior to implantation (CL and endometrium), both of which are critical to reproductive success. Important results, study limitations, and some ideas for future research have been outlined in the following sections.

## 5.1 Key findings

A comprehensive list of genes and miRNAs expressed in three tissues vital to reproduction was determined, and this list is much longer than the previously generated lists. Most RNA-seq-based studies have focused on either one breed of sheep or a single tissue (see 1.4.5). The unique materials (including tissues, sequence data, and phenotype records) used in this thesis can extend or enrich the findings of previous studies. The pilot study compiled the list of genes and miRNAs expressed in the Finnsheep ovaries. The flushing diet appeared to influence reproductive performance in low-prolific Texel sheep, mainly during ovulation, and, to some extent, embryo viability. Furthermore, despite sharing 50% of the genome each from Finnsheep and Texel sheep, the F1 crosses were closer to Finnsheep in terms of gene expression, possibly owing to POEs such as genomic imprinting. Therefore, the gene expression profiles of F1 crosses and the prevalence of non-Mendelian inheritance should be further examined using more samples. These results of the diet experiment and F1 crosses can be used to implement sheep breeding strategies aiming for optimum reproductive capacity.

The second phase of the main experiment sheds light on the dynamics of gene expression during early pregnancy in sheep, mediated by the CL and endometrium. The shared and tissue-specific roles of the CL and endometrium were evident by the commonly expressed genes and DEGs, respectively. The importance of the immune system during early pregnancy was indicated by the gene expression profile: the immune-related genes were upregulated in Finnsheep CL compared to their expression in Texel sheep CL. Similarly, differential expression of *TXNL1* (upregulated in Finnsheep and F1 crosses, compared to Texel) indicated potential for embryonic diapause in sheep. After day 12 of pregnancy, the gene expression in the endometrium appeared to change remarkably, whereas it remained more or less the same in the CL. One important phenotypic observation made from the second phase of the experiment was that Finnsheep had 4.1 CLs on average, indicating a much higher ovulation rate compared to the number of embryos formed (2.8 on average). This indicates that the reproductive efficiency of Finnsheep can still be improved. Finally, the comparative transcriptome profiling of the two reproductive tissues (ovary and endometrium) revealed that the reproductive characteristics of Finnsheep were superior to those of typically monotocous European mouflons.

The number of miRNAs (>500) quantified in this thesis is more than three-folds higher than that available in the miRBase database (n = 153). A notable finding from the miRNA-Seq data was the identification and characterization of maternally imprinted sheep C18MC, which included 46 miRNAs. Further studies involving the influence of ERVs and C18MC on sheep reproduction are compelling.

## 5.2 Study limitations

Gene expression analysis in the endometrium and CL may have been impacted by the flushing diet. The same ewes that lost one of their ovaries during the first phase of the experiment were used for collecting CL biopsy samples. Therefore, the absence of one of the ovaries and the possible negative feedback effect may have influenced the gene expression profiles of the CL and endometrium. Additionally, biopsy samples show tissue heterogeneity; thus, future experiments based on certain cell types may be useful. Laser capture microdissection, for example, can be used. The developmental progression in the embryos, particularly in the endometrial tissues, appeared to affect the overall gene expression.

The results obtained by comparing Finnsheep and European mouflon sheep may have been biased by the age of the mouflon females, stage of the estrous cycle at the time of tissue collection, etc. Further, the age at sexual maturity of Finnsheep and European mouflon sheep is 6 months (Maijala and Österberg, 1977) and 1.5 years (Cugnasse et al., 1985), respectively, and the differences in the sexual maturation age may have led to variations in gene expression. The quality of the sheep reference genome and annotation may have also affected the

results. On average, ∼50% of transcripts lack gene annotation and ∼60% lack functional (GO and KEGG) annotations; thus, the current version of the sheep reference genome (Oar 3.1) includes complete annotations at the functional level for only approximately 40% of genes. In many cases, homology-based approaches (e.g., BLAST) have been used; however, the results may not always be accurate. Moreover, some genes could not be discovered using BLAST, and these genes have a higher probability of being associated with species-specific or even breed-specific characteristics.

## 5.3 Future directions

Evaluation using more breeds, tissue samples (involving reproductive and immune systems), study points (reproduction stages), and deeper sequencing would provide more in-depth results. Adopting novel techniques such as single-cell RNA seq can provide cell-specific gene expression resolution, whereas applying integrative omics and spatial transcriptomics can provide a more holistic view. Furthermore, the newly hypothesized omnigenic model (Boyle et al., 2017; Liu et al., 2019) could also be applied for studying complex fertility traits. While the primary focus of this thesis was prolificacy, this is not the only important phenotype. Sheep have multiple uses, and traits associated with other functions such as milk, meat, and wool production are also of high economic importance.

The identification of genes relevant to POEs, including imprinted genes in each tissue evaluated in this thesis and additionally in the embryos (also collected but not yet sequenced), may reveal the heritability of important genes. Imprinted genes are known to be regulated by lncRNAs and are typically close to each other in a group (Barlow and Bartolomei, 2014). The high coverage (∼80 million reads per sample) of the sequence data allows for both lncRNA and allele-specific expression analyses of F1 crosses. Future experiment assessing the effect of GDF9 (V371M) mutation in overall gene expression is compelling whereby animals carrying the mutation could be compared against those lacking it.

To study the role of sheep C18MC, epigenetic analyses in F1 crosses can be performed by deleting the imprinted region DLK1-DIO3 upstream of C18MC (using genome editing technologies such as CRISPR-Cas9 system) in the paternally- and the maternally-inherited chromosomes of the embryos. Next, the expression of the miRNA cluster can be identified. The deletion of the methylated region in the maternal chromosome is expected to suppress the expression of the miRNA cluster, whereas the deletion of the same region in the paternal chromosome should not change the expression levels of these miRNAs.

In addition to the major role of genetics, other factors are also important. The size of the uterus is essential for accommodating multiple fetal lambs. Farm management practices, including care for additional lambs, reportedly play a sig-

nificant role in preventing lamb mortality. Although we observed that the flushing diet might improve the ovulation rate, the effect may not necessarily be realized in the number of live births, as observed in this thesis. Therefore, time-series experiments involving folliculogenesis, fertilization, and further developments may provide insights in terms of embryo mortality. Moreover, the possibility of embryonic diapause and its link with out-of-season breeding in Finnsheep could be an interesting area of research.

The technological aspects of genomics have vigorously advanced in recent years. For instance, RNA seq has quickly become more widely used than microarray-based methods, and new single-cell sequencing methods and spatial transcriptomics will soon become the standard choices. In contrast, the tools and resources developed for evaluating sequence data are constantly evolving. Some new tools have been developed, some previously developed tools have been updated, and others are no longer in use or have been replaced. The data generated as part of this thesis is unique and will be useful for both follow-up studies and projects dependent on publicly available datasets (e.g., FAANG). The quality of the sheep reference genome and annotations are expected to improve in the near future, and a new Rambouillet sheep reference genome is now available in Ensembl. We are currently sequencing the genome of Finnsheep, which is expected to be completed in 2021, with full annotation. Analyzing the dataset with this new resource will generate the information that may have been currently missed in this thesis. It would also be interesting to determine the overall gene expression dynamics of the embryos, which may provide further insights into the crosstalk between the CL, endometrium, and embryo. Validation experiments of candidate genes and regions using gene knockout and genome editing technologies will support the findings of this thesis.

Native sheep breeds that have survived several breed-related demographic and even societal drifts (e.g., winter war) and thrived in harsh climatic conditions, such as those in Finland, are genetic resources with high socioeconomic importance. The continuous effort of the ISGC, FAANG, and several research groups have contributed significantly to sheep research. The next challenge is the preservation and sustainable utilization of these (with special emphasis on marginalized breeds) valuable animal genetic resources. Sequencing costs have dramatically decreased, and generating a nearly complete annotated genome of a species in one project is possible. Therefore, producing breed-specific reference genomes as a part of national conservation programs is recommended for better characterization of important traits and even the whole organism.

# Acknowledgements

This dissertation was conducted at the Natural Resources Institute Finland (Luke), Jokioinen, and the University of Eastern Finland, Kuopio, between 2013 and 2020. The funding for this study was provided by the Academy of Finland, Luke, the University of Helsinki and the Niemi Foundation. I acknowledge the CSC – IT Center for Science, Finland, for providing computational resources.

My warmest gratitude goes to my supervisors, Professors Juha Kantanen and Meng-Hua Li. Professor Kantanen provided constant support and sound advice throughout my graduate studies. I warmly thank my thesis committee members, Professor Päivi Onkamo and Dr. Eija Korpelainen, for their support and guidance.

I appreciate the valuable feedback on my dissertation manuscript from my pre-examiners Dr. Stephane Fabre and Professor Thomas Spencer.

I thank all my co-authors for their valuable contribution to the publications. Not to forget the editors and anonymous reviewers that have helped in improving the quality of the published articles. Many thanks to Dr. Jaana Peippo for helping me interpret the results and fruitful discussions. Thank you Dr. Xiaoju Hu and Dr. Yang Ji for your contributions to the publications II and VII.

I am deeply indebted to all my colleagues and collaborators and especially to Mervi Honkatukia, Johanna Rautiainen, Tuula-Marjatta Hamama, Anu Tuomola, Arja Seppälä, Kati Kaisajoki, Ilma Tapio, Anneli Virta, Mervi Mutikainen, Annukka Numminen, and Kalle Saastamoinen for their valuable expertise and assistance.

Heartfelt thanks go to all my colleagues at Luke and UEF for their supportive attitude and friendliness. Melak, it was a pleasure sharing an office with you; thanks for being such a nice friend.

I am deeply thankful to my wife for her unconditional love and support and to our son for always putting a smile on my face.

Jokioinen, October 2020

# References


Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4:e05005.

Ali, A. T., Boehme, L., Carbajosa, G., Seitan, V. C., Small, K. S., and Hodgkinson, A. (2019). Nuclear genetic regulation of the human mitochondrial transcriptome. *eLife*, 8:e41927.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–9.

Andrés-León, E., Núñez-Torres, R., and Rojas, A. M. (2016). miARma-Seq: A comprehensive tool for miRNA, mRNA and circRNA analysis. *Scientific Reports*, 6:25749.

Andrews, S. (2019). A quality control analysis tool for high throughput sequencing data: https://github.com/s-andrews/fastqc.

Armstrong, E., Iriarte, A., Nicolini, P., Santos, J. D. L., Ithurralde, J., Bielli, A., Bianchi, G., and Peñagaricano, F. (2018). Comparison of transcriptomic landscapes of different lamb muscles using RNA-Seq. *PLoS ONE*, 13(7):e0200732.

Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., Golder, M., Mura, M., Zhang, Y.-p., Yu, L., Pereira, F., DeMartini, J. C., Leymaster, K., Spencer, T. E., and Palmarini, M. (2007). A paradigm for virus–host coevolution: Sequential counter-adaptations between endogenous and exogenous retroviruses. *PLOS Pathogens*, 3(11):e170.

Avella, M., Baibakov, B., Sadusky, A. B., and Dean, J. (2016). ZP2 peptide beads that select human sperm in vitro, decoy mouse sperm in vivo, and provide reversible contraception. *Fertility and Sterility*, 106(3):e112.

Baird, D. T. (1987). A model for follicular selection and ovulation: Lessons from superovulation. *Journal of Steroid Biochemistry*, 27(1-3):15–23.

Baird, D. T. and Campbell, B. K. (1998). Follicle selection in sheep with breed differences in ovulation rate. *Molecular and Cellular Endocrinology*, 145(1):89–95.

Bakhtiarizadeh, M. R., Salehi, A., Alamouti, A. A., Abdollahi-Arpanahi, R., and Salami, S. A. (2019). Deep transcriptome analysis using RNA-Seq suggests

novel insights into molecular aspects of fat-tail metabolism in sheep. *Scientific Reports*, 9:9203.

Barlow, D. P. and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harbor Perspectives in Biology*, 4(6):a018382.

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–97.

Bartlewski, P. M., Baby, T. E., and Giffin, J. L. (2011). Reproductive cycles in sheep. *Animal Reproduction Science*, 124(3-4):259–68.

Bartlewski, P. M., Beard, A. P., Cook, S. J., Chandolia, R. K., Honaramooz, A., and Rawlings, N. C. (1999). Ovarian antral follicular dynamics and their relationships with endocrine variables throughout the oestrous cycle in breeds of sheep differing in prolificacy. *Reproduction*, 115(1):111–24.

Bazer, F. W. (2013). Pregnancy recognition signaling mechanisms in ruminants and pigs. *Journal of Animal Science and Biotechnology*, 4(1):23.

Ben Jemaa, S., Ruesche, J., Sarry, J., Woloszyn, F., Lassoued, N., and Fabre, S. (2019). The high prolificacy of D'man sheep is associated with the segregation of the FecLL mutation in the B4GALNT2 gene. *Reproduction in Domestic Animals*, 54(3):531–37.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y., and Bentwich, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genetics*, 37(7):766–70.

Bermejo, L. A., Mellado, M., Mata, J., Arévalo, J. R., Nascimento, L. D., and Camacho, A. (2010). Factors affecting days to conception, litter size and litter weight of intensively managed Canarian hair sheep. *Journal of Applied Animal Research*, 37(2):261–4.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8):1091–3.

Bodin, L., Di Pasquale, E., Fabre, S., Bontoux, M., Monget, P., Persani, L., and Mulsant, P. (2007). A novel mutation in the bone morphogenetic protein 15 gene causing defective protein secretion is associated with both increased ovulation rate and sterility in Lacaune sheep. *Endocrinology*, 148(1):393–400.

Bolet, G. (1986). Timing and extent of embryonic mortality in pigs, sheep and goats: Genetic variability. In Sreenan, J. M. and Diskin, M. G., editors, *Embryonic Mortality in Farm Animals*, Current Topics in Veterinary Medicine and Animal Science, pages 12–43. Springer Netherlands, Dordrecht.

Bonnet, A., Bevilacqua, C., Benne, F., Bodin, L., Cotinot, C., Liaubet, L., Sancristobal, M., Sarry, J., Terenina, E., Martin, P., Tosser-Klopp, G., and Mandon-Pepin, B. (2011). Transcriptome profiling of sheep granulosa cells and oocytes during early follicular development obtained by laser capture microdissection.

*BMC Genomics*, 12:417.

Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7):1177–86.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–7.

Brooks, K. (2016). *Biological Role of Conceptus Derived Factors during Early Pregnancy in Ruminants*. Thesis, University of Missouri–Columbia.

Brooks, K., Burns, G., and Spencer, T. E. (2014). Conceptus elongation in ruminants: Roles of progesterone, prostaglandin, interferon tau and cortisol. *Journal of Animal Science and Biotechnology*, 5(1):53.

Brooks, K., Burns, G. W., Moraes, J. G. N., and Spencer, T. E. (2016). Analysis of the uterine epithelial and conceptus transcriptome and luminal fluid proteome during the peri-implantation period of pregnancy in sheep. *Biology of Reproduction*, 95(4):88.

Brooks, K. and Spencer, T. E. (2015). Biological roles of interferon tau (IFN-T) and type I IFN receptors in elongation of the ovine conceptus. *Biology of Reproduction*, 92(2):47.

Burkart, A. D., Xiong, B., Baibakov, B., Jiménez-Movilla, M., and Dean, J. (2012). Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy. *Journal of Cell Biology*, 197(1):37–44.

Chamberlain, A. J., Vander Jagt, C. J., Hayes, B. J., Khansefid, M., Marett, L. C., Millen, C. A., Nguyen, T. T. T., and Goddard, M. E. (2015). Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics*, 16:993.

Chaturvedi, O. H., Bhatta, R., Verma, D. L., and Singh, N. P. (2006). Effect of flushing on nutrient utilization and reproductive performance of ewes grazing on community rangeland. *Asian-Australasian Journal of Animal Sciences*, 19(4):521–5.

Chen, H. Y., Shen, H., Jia, B., Zhang, Y. S., Wang, X. H., and Zeng, X. C. (2015). Differential gene expression in ovaries of Qira Black sheep and Hetian sheep using RNA-seq technique. *PLoS ONE*, 10(3):e0120170.

Chen, L., Liu, K., Zhao, Z., Blair, H. T., Zhang, P., Li, D., and Ma, R. Z. (2012). Identification of sheep ovary genes potentially associated with off-season reproduction. *Journal of Genetics and Genomics*, 39(4):181–90.

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R. R., Pemberton, J. M., Beraldi, D., Stear, M. J., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M. H., Kazwala, R. R., Zhang, Y.-P., Arranz, J. J., Ali, B. A., Wang, Z., Uzun, M., Dione, M. M., Olsaker, I., Holm, L.-E., Saarma, U., Ahmad, S., Marzanov, N., Eythorsdottir, E., Holland, M. J., Ajmone-Marsan, P., Bruford, M. W., Kantanen, J., Spencer, T. E., and Palmarini, M. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science*, 324(5926):532–6.

Chopra-Dewasthaly, R., Korb, M., Brunthaler, R., and Ertl, R. (2017). Comprehensive RNA-seq profiling to evaluate the sheep mammary gland transcriptome in response to experimental Mycoplasma agalactiae infection. *PLoS ONE*, 12(1):e0170015.

Chu, M. X., Liu, Z. H., Jiao, C. L., He, Y. Q., Fang, L., Ye, S. C., Chen, G. H., and Wang, J. Y. (2007). Mutations in BMPR-IB and BMP-15 genes are associated with litter size in Small Tailed Han sheep (Ovis aries). *Journal of Animal Science*, 85(3):598–603.

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6):1767–71.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18):3674–6.

Coster, A., Madsen, O., Heuven, H. C. M., Dibbits, B., Groenen, M. A. M., van Arendonk, J. A. M., and Bovenhuis, H. (2012). The imprinted gene DIO3 is a candidate gene for litter size in pigs. *PLoS ONE*, 7(2):e31825.

Crocker, P. R., Paulson, J. C., and Varki, A. (2007). Siglecs and their roles in the immune system. *Nature Reviews Immunology*, 7(4):255–66.

Cugnasse, J. M., García, M., and Veyrac, T. (1985). Contribution a l' etude de la biologie de la reproduction du mouflon (Ovis ammon musimon), par examens post-mortem, dans le massif du Caroux-Espinouse. *Bulletin mensuel ONC*, 89:33–5.

da Rocha, S. T., Edwards, C. A., Ito, M., Ogata, T., and Ferguson-Smith, A. C. (2008). Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends in Genetics*, 24(6):306–16.

Davis, G. H. (2005). Major genes affecting ovulation rate in sheep. *Genetics, selection, evolution : GSE*, 37 Suppl 1:S11–23.

Davis, G. H., Balakrishnan, L., Ross, I. K., Wilson, T., Galloway, S. M., Lumsden, B. M., Hanrahan, J. P., Mullen, M., Mao, X. Z., Wang, G. L., Zhao, Z. S., Zeng, Y. Q., Robinson, J. J., Mavrogenis, A. P., Papachristoforou, C., Peter, C., Baumung, R., Cardyn, P., Boujenane, I., Cockett, N. E., Eythorsdottir, E., Arranz, J. J., and Notter, D. R. (2006). Investigation of the Booroola (FecB) and Inverdale (FecX(I)) mutations in 21 prolific breeds and strains of sheep sampled in 13 countries. *Animal Reproduction Science*, 92(1-2):87–96.

Davis, G. H., Dodds, K. G., Wheeler, R., and Jay, N. P. (2001). Evidence that an imprinted gene on the X chromosome increases ovulation rate in sheep. *Biology of Reproduction*, 64(1):216–21.

Davis, G. H., Galloway, S. M., Ross, I. K., Gregan, S. M., Ward, J., Nimbkar, B. V., Ghalsasi, P. M., Nimbkar, C., Gray, G. D., Subandriyo, Inounu, I., Tiesnamurti, B., Martyniuk, E., Eythorsdottir, E., Mulsant, P., Lecerf, F., Hanrahan, J. P., Bradford, G. E., and Wilson, T. (2002). DNA tests in prolific sheep from

eight countries provide new evidence on origin of the Booroola (FecB) mutation. *Biology of Reproduction*, 66(6):1869–74.

Davis, J. S. and LaVoie, H. A. (2018). *Molecular Regulation of Progesterone Production in the Corpus Luteum*. Elsevier Inc., 3 edition.

Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., Martyniuk, E., and Bodin, L. (2013). Genome-wide association studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype in sheep. *PLoS Genetics*, 9(4):e1003482.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, 4(9):R60.

Dillon, P., Crosse, S., O'Brien, B., and Mayes, R. W. (2002). The effect of forage type and level of concentrate supplementation on the performance of spring-calving dairy cows in early lactation. *Grass and Forage Science*, 57(3):212–23.

Dini, P., Daels, P., Loux, S. C., Esteller-Vico, A., Carossino, M., Scoggin, K. E., and Ball, B. A. (2018). Kinetics of the chromosome 14 microRNA cluster ortholog and its potential role during placental development in the pregnant mare. *BMC Genomics*, 19:954.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

Drouilhet, L., Mansanet, C., Sarry, J., Tabet, K., Bardou, P., Woloszyn, F., Lluch, J., Harichaux, G., Viguié, C., Monniaux, D., Bodin, L., Mulsant, P., and Fabre, S. (2013). The highly prolific phenotype of Lacaune sheep is associated with an ectopic expression of the B4GALNT2 gene within the ovary. *PLoS Genetics*, 9(9):e1003809.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, 38:W64–70.

Dunlap, K. A., Palmarini, M., Varela, M., Burghardt, R. C., Hayashi, K., Farmer, J. L., and Spencer, T. E. (2006). Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(39):14390–5.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8):1184–91.

Everaert, C., Luypaert, M., Maag, J. L. V., Cheng, Q. X., Dinger, M. E., Hellemans, J., and Mestdagh, P. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Scientific Reports*, 7:1559.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*,

32(19):3047–8.

Fabre, S., Pierre, A., Mulsant, P., Bodin, L., Di Pasquale, E., Persani, L., Monget, P., and Monniaux, D. (2006). Regulation of ovulation rate in mammals: Contribution of sheep genetic models. *Reproductive Biology and Endocrinology*, 4(1):20.

Fahmy, M. H. and Dufour, J. J. (1988). The accumulative effect of Finnsheep breeding in crossbreeding schemes: Reproductive performance. *Canadian Journal of Animal Science*, 68(1):69–81.

Farin, C. E., Imakawa, K., Hansen, T. R., McDonnell, J. J., Murphy, C. N., Farin, P. W., and Roberts, R. M. (1990). Expression of trophoblastic interferon genes in sheep and cattle. *Biology of Reproduction*, 43(2):210–8.

Farin, C. E., Sawyer, H. R., and Niswender, G. D. (1989). Analysis of cell types in the corpus luteum of the sheep. *Journal of Reproduction and Fertility. Supplement*, 37:181–7.

Feng, X., Li, F., Wang, F., Zhang, G., Pang, J., Ren, C., Zhang, T., Yang, H., Wang, Z., and Zhang, Y. (2018). Genome-wide differential expression profiling of mRNAs and lncRNAs associated with prolificacy in Hu sheep. *Bioscience Reports*, 38(2):BSR20171350.

Filant, J. and Spencer, T. E. (2014). Uterine glands: Biological roles in conceptus implantation, uterine receptivity, and decidualization. *The Internation Journal of Developmental Biology*, 58(0):107–16.

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, 40(1):37–52.

Friedley, N. J. and Rosen, S. (1975). Carbonic anhydrase activity in the mammalian ovary, fallopian tube, and uterus: Histochemical and biochemical studies. *Biology of Reproduction*, 12(2):293–304.

Galloway, S. M., McNatty, K. P., Cambridge, L. M., Laitinen, M. P., Juengel, J. L., Jokiranta, T. S., McLaren, R. J., Luiro, K., Dodds, K. G., Montgomery, G. W., Beattie, A. E., Davis, G. H., and Ritvos, O. (2000). Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature genetics*, 25(3):279–83.

Ginther, O. J., Kot, K., and Wiltbank, M. C. (1995). Associations between emergence of follicular waves and fluctuations in FSH concentrations during the estrous cycle in ewes. *Theriogenology*, 43(3):689–703.

Glazov, E. A., McWilliam, S., Barris, W. C., and Dalrymple, B. P. (2008). Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Molecular Biology and Evolution*, 25(5):939–48.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., and Palma, F. (2011). Full-length

transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7):644–52.

Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008). miRBase: Tools for microRNA genomics. *Nucleic Acids Research*, 36:D154–D158.

Guerrini, M., Forcina, G., Panayides, P., Lorenzini, R., Garel, M., Anayiotos, P., Kassinis, N., and Barbanera, F. (2015). Molecular DNA identity of the mouflon of Cyprus (Ovis orientalis ophion, Bovidae): Near Eastern origin and divergence from Western Mediterranean conspecific populations. *Systematics and Biodiversity*, 13(5):472–83.

Guillocheau, G. M., Hou, A. E., Meersseman, C., Esquerré, D., Rebours, E., Letaief, R., Simao, M., Hypolite, N., Bourneuf, E., Bruneau, N., Vaiman, A., Jagt, C. J. V., Chamberlain, A. J., and Rocha, D. (2019). Survey of allele specific expression in bovine muscle. *Scientific Reports*, 9:4297.

Hanrahan, J. P., Gregan, S. M., Mulsant, P., Mullen, M., Davis, G. H., Powell, R., and Galloway, S. M. (2004). Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (Ovis aries). *Biology of Reproduction*, 70(4):900–9.

Harris, R. S. (2007). *Improved Pairwise Alignment of Genomic DNA*. Ph.D. Thesis, The Pennsylvania State University.

Hernandez-Medrano, J. H., Campbell, B. K., and Webb, R. (2012). Nutritional influences on folliculogenesis. *Reproduction in Domestic Animals*, 47 Suppl 4:274–82.

Hernández-Montiel, W., Collí-Dula, R. C., Ramón-Ugalde, J. P., Martínez-Núñez, M. A., and Zamora-Bustillos, R. (2019). RNA-seq transcriptome analysis in ovarian tissue of Pelibuey breed to explore the regulation of prolificacy. *Genes*, 10(5):358.

Hiendleder, S., Kaupe, B., Wassmuth, R., and Janke, A. (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1494):893–904.

Hrdlickova, R., Toloue, M., and Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews. RNA*, 8(1).

Hynninen, P., Hämäläinen, J. M., Pastorekova, S., Pastorek, J., Waheed, A., Sly, W. S., Tomas, E., Kirkinen, P., and Parkkila, S. (2004). Transmembrane carbonic anhydrase isozymes IX and XII in the female mouse reproductive organs. *Reproductive Biology and Endocrinology*, 2:73.

Jäger, M., Ott, C.-E., Grünhagen, J., Hecht, J., Schell, H., Mundlos, S., Duda, G. N., Robinson, P. N., and Lienau, J. (2011). Composite transcriptome assembly of RNA-seq data in a sheep model for delayed bone healing. *BMC Genomics*, 12:158.

Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C.,

Muzny, D. M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Fuentes-Utrilla, P., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L., Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristiansen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J. W., Wang, J., Worley, K. C., Archibald, A. L., Cockett, N., Xu, X., Wang, W., and Dalrymple, B. P. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188):1168–73.

Kantanen, J., Løvendahl, P., Strandberg, E., Eythorsdottir, E., Li, M.-H., Kettunen, A., Berg, P., and Meuwissen, T. (2015). Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. *Frontiers in Genetics*, 6:52.

Karagianni, A. E., Vasoya, D., Finlayson, J., Martineau, H. M., Wood, A. R., Cousens, C., Dagleish, M. P., Watson, M., and Griffiths, D. J. (2019). Transcriptional response of ovine lung to infection with jaagsiekte sheep retrovirus. *Journal of Virology*, 93:e00876–19.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–60.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36.

Kim, Y.-K. and Kim, V. N. (2007). Processing of intronic microRNAs. *EMBO J*, 26(3):775–83.

Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M. J. (2013). Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology*, 14(3):153–65.

Krueger, F. (2019). A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data: https://github.com/felixkrueger/trimgalore.

Kumar, S., Mishra, A. K., Kolte, A. P., Dash, S. K., and Karim, S. A. (2008). Screening for Booroola (FecB) and Galway (FecXG) mutations in Indian sheep. *Small Ruminant Research*, 80(1):57–61.

La, Y., Tang, J., He, X., Di, R., Wang, X., Liu, Q., Zhang, L., Zhang, X., Zhang, J., Hu, W., and Chu, M. (2019). Identification and characterization of mRNAs and lncRNAs in the uterus of polytocous and monotocous Small Tail Han sheep (Ovis aries). *PeerJ*, 7:e6938.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.

*Genome Biology*, 10(3):R25.

Lassoued, N., Benkhlil, Z., Woloszyn, F., Rejeb, A., Aouina, M., Rekik, M., Fabre, S., and Bedhiaf-Romdhani, S. (2017). FecXBar a Novel BMP15 mutation responsible for prolificacy and female sterility in Tunisian Barbarine Sheep. *BMC Genetics*, 18:43.

Lawson, H. A., Cheverud, J. M., and Wolf, J. B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. *Nature Reviews Genetics*, 14(9):609–17.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–54.

Li, B. and Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12:323.

Li, C.-Y., Li, X., Liu, Z., Ni, W., Zhang, X., Hazi, W., Ma, Q., Zhang, Y., Cao, Y., Qi, J., Yao, Y., Feng, L., Wang, D., Hou, X., Yu, S., Liu, L., Zhang, M., and Hu, S. (2019). Identification and characterization of long non-coding RNA in prenatal and postnatal skeletal muscle of sheep. *Genomics*, 111(2):133–41.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–9.

Liu, X., Li, Y. I., and Pritchard, J. K. (2019). Trans effects on gene expression can drive omnigenic inheritance. *Cell*, 177(4):1022–34.e6.

Liu, X.-M., Ding, G.-L., Jiang, Y., Pan, H.-J., Zhang, D., Wang, T.-T., Zhang, R.-J., Shu, J., Sheng, J.-Z., and Huang, H.-F. (2012). Down-regulation of S100A11, a calcium-binding protein, in human endometrium may cause reproductive failure. *Journal of Clinical Endocrinology and Metabolism*, 97(10):3672–83.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.

Ma, L., Zhang, M., Jin, Y., Erdenee, S., Hu, L., Chen, H., Cai, Y., and Lan, X. (2018). Comparative transcriptome profiling of mRNA and lncRNA related to tail adipose tissues of sheep. *Frontiers in Genetics*, 9:365.

Madden, T. (2003). *The BLAST Sequence Analysis Tool*. National Center for Biotechnology Information (US).

Maijala, K. (1988). History, recent development and uses of Finnsheep. *Journal of Agricultural Science Finland*, 60:449–54.

Maijala, K. and Österberg, S. (1977). Productivity of pure Finnsheep in Finland and abroad. *Livestock Production Science*, 4(4):355–77.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput

sequencing reads. *EMBnet.journal*, 17(1):10–12.

Martinez-Royo, A., Dervishi, E., Alabart, J. L., Jurado, J. J., Folch, J., and Calvo, J. H. (2009). Freemartinism and FecXR allele determination in replacement ewes of the Rasa Aragonesa sheep breed by duplex PCR. *Theriogenology*, 72(8):1148–52.

McBride, D., Carré, W., Sontakke, S. D., Hogg, C. O., Law, A., Donadeu, F. X., and Clinton, M. (2012). Identification of miRNAs associated with the follicular-luteal transition in the ruminant ovary. *Reproduction*, 144(2):221–33.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–303.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1):122.

Mercer, T. R., Neph, S., Dinger, M. E., Crawford, J., Smith, M. A., Shearwood, A.-M. J., Haugen, E., Bracken, C. P., Rackham, O., Stamatoyannopoulos, J. A., Filipovska, A., and Mattick, J. S. (2011). The human mitochondrial transcriptome. *Cell*, 146(4):645–58.

Mesquita, F. S., Ramos, R. S., Pugliesi, G., Andrade, S. C. S., Van Hoeck, V., Langbeen, A., Oliveira, M. L., Gonella-Diaza, A. M., Gasparin, G., Fukumasu, H., Pulz, L. H., Membrive, C. M., Coutinho, L. L., and Binelli, M. (2015). The receptive endometrial transcriptomic signature indicates an earlier shift from proliferation to metabolism at early diestrus in the cow. *Biology of Reproduction*, 93(2):52.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2016). PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Research*, 44(D1):D336–D342.

Miao, X. and Luo, Q. (2013). Genome-wide transcriptome analysis between small-tail Han sheep and the Surabaya fur sheep using high-throughput RNA sequencing. *Reproduction*, 145(6):587–96.

Minten, M. A., Bilby, T. R., Bruno, R. G. S., Allen, C. C., Madsen, C. A., Wang, Z., Sawyer, J. E., Tibary, A., Neibergs, H. L., Geary, T. W., Bauersachs, S., and Spencer, T. E. (2013). Effects of fertility on gene expression and function of the bovine endometrium. *PLoS ONE*, 8(8):e69444.

Monniaux, D., Clément, F., Dalbiès-Tran, R., Estienne, A., Fabre, S., Mansanet, C., and Monget, P. (2014). The ovarian reserve of primordial follicles and the dynamic reserve of antral growing follicles: What is the link? *Biology of Reproduction*, 90(4):85.

Moradband, F., Rahimi, G., and Gholizadeh, M. (2011). Association of polymorphisms in fecundity genes of GDF9, BMP15 and BMP15-1B with litter size in Iranian Baluchi sheep. *Asian-Australasian Journal of Animal Sciences*,

24(9):1179–83.

Morris, K. V. and Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, 15(6):423–37.

Mott, R., Yuan, W., Kaisaki, P., Gan, X., Cleak, J., Edwards, A., Baud, A., and Flint, J. (2014). The architecture of parent-of-origin effects in mice. *Cell*, 156(1):332–42.

Mullen, M. P. and Hanrahan, J. P. (2014). Direct evidence on the contribution of a missense mutation in GDF9 to variation in ovulation rate of Finnsheep. *PLoS ONE*, 9(4):e95251.

Mullen, M. P., Hanrahan, J. P., Howard, D. J., and Powell, R. (2013). Investigation of prolific sheep from UK and Ireland for evidence on origin of the mutations in BMP15 (FecX(G), FecX(B)) and GDF9 (FecG(H)) in Belclare and Cambridge sheep. *PLoS ONE*, 8(1):e53172.

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., Pisselet, C., Riquet, J., Monniaux, D., Callebaut, I., Cribiu, E., Thimonier, J., Teyssier, J., Bodin, L., Cognié, Y., Chitour, N., and Elsen, J. M. (2001). Mutation in Bone Morphogenetic Protein Receptor-IB is associated with increased ovulation rate in Booroola Mérino ewes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9):5104–9.

Nicol, L., Bishop, S. C., Pong-Wong, R., Bendixen, C., Holm, L.-E., Rhind, S. M., and McNeilly, A. S. (2009). Homozygosity for a single base-pair mutation in the oocyte-specific GDF9 gene results in sterility in Thoka sheep. *Reproduction*, 138(6):921–33.

Noguer-Dance, M., Abu-Amero, S., Al-Khtib, M., Lefèvre, A., Coullin, P., Moore, G. E., and Cavaillé, J. (2010). The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Human Molecular Genetics*, 19(18):3566–82.

Nosrati, M., Nanaei, H. A., Ghanatsaman, Z. A., and Esmailizadeh, A. (2019). Whole genome sequence analysis to detect signatures of positive selection for high fecundity in sheep. *Reproduction in Domestic Animals*, 54(2):358–64.

Notter, D. R. (2008). Genetic aspects of reproduction in sheep. *Reproduction in Domestic Animals*, 43 Suppl 2:122–8.

Notter, n. (2000). Effects of ewe age and season of lambing on prolificacy in US Targhee, Suffolk, and Polypay sheep. *Small Ruminant Research*, 38(1):1–7.

Oreper, D., Schoenrock, S. A., McMullan, R., Ervin, R., Farrington, J., Miller, D. R., de Villena, F. P.-M., Valdar, W., and Tarantino, L. M. (2018). Reciprocal F1 hybrids of two inbred mouse strains reveal parent-of-origin and perinatal diet effects on behavior and expression. *G3: Genes, Genomes, Genetics*, 8(11):3447–68.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon: Fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature Methods*, 14(4):417–9.

Pedrosa, S., Uzun, M., Arranz, J.-J., Gutiérrez-Gil, B., San Primitivo, F., and Bayón, Y. (2005). Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society B: Biological Sciences*, 272(1577):2211–7.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3):290–5.

Petrovic, M., Caro-Petrovic, V., Ruzic-Muslic, D., Maksimovic, N., Ilic, Z., Milosevic, B., and Stojkovic, J. (2012). Some important factors affecting fertility in sheep. *Biotechnology in Animal Husbandry*, 28(3):517–28.

Pillai, S., Netravali, I. A., Cariappa, A., and Mattoo, H. (2012). Siglecs and immune regulation. *Annual Review of Immunology*, 30:357–92.

Poplin, F. (1979). Origine du Mouflon de Corse dans une nouvelle perspective paléontologique: Par marronnage. *Annales de Génétique et de Sélection Animale*, 11(2):133.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–75.

Quinlivan, T. D., Martin, C. A., Taylor, W. B., and Cairney, I. M. (1966). Estimates of pre- and perinatal mortality in the New Zealand Romney Marsh ewe. II. Pre- and perinatal loss in those ewes that conceived to second service and those that returned to second service and were mated a third time. *Journal of Reproduction and Fertility*, 11(3):391–8.

Ramsköld, D., Wang, E. T., Burge, C. B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology*, 5(12):e1000598.

Rhind, S. M., Robinson, J. J., Fraser, C., and McHattie, I. (1980). Ovulation and embryo survival rates and plasma progesterone concentrations of prolific ewes treated with PMSG. *Journal of Reproduction and Fertility*, 58(1):139–44.

Ricordeau, G., Thimonier, J., Poivey, J., Driancourt, M., Hochereau-De-Reviers, M., and Tchamitchian, L. (1990). I.N.R.A. research on the Romanov sheep breed in France: A review. *Livestock Production Science*, 24(4):305–32.

Robinson, J. (1996). Nutrition and reproduction. *Animal Reproduction Science*, 42(1-4):25–34.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–40.

Ruusila, V. and Kojola, I. (2010). *Ungulates and Their Management in Finland*. Cambridge University Press.

Ryder, M. L. (1981). A survey of European primitive breeds of sheep. *Annales de génétique et de sélection animale*, 13(4):381.

Ryder, M. L. (1984). Sheep. In Mason, I. L., editor, *Evolution of Domestic Animals*, pages 63–85. Longman, London.

Sahraeian, S. M. E., Mohiyuddin, M., Sebra, R., Tilgner, H., Afshar, P. T., Au, K. F., Asadi, N. B., Gerstein, M. B., Wong, W. H., Snyder, M. P., Schadt, E., and Lam, H. Y. K. (2017). Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nature Communications*, 8:59.

Salavati, M., Bush, S. J., Palma-Vera, S., McCulloch, M. E. B., Hume, D. A., and Clark, E. L. (2019). Elimination of reference mapping bias reveals robust immune related allele-specific expression in crossbred sheep. *Frontiers in Genetics*, 10:863.

Sangiovanni, M., Granata, I., Thind, A. S., and Guarracino, M. R. (2019). From trash to treasure: Detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics*, 20:168.

Sarda, I. R., Robertson, H. A., and Smeaton, T. C. (1973). Sequential changes in plasma progesterone levels in the ewe during the estrous cycle, and during pregnancy in intact and ovariectomized sheep. *Canadian Journal of Animal Science*, 53(1):25–34.

Scaramuzzi, R. J., Adams, N. R., Baird, D. T., Campbell, B. K., Downing, J. A., Findlay, J. K., Henderson, K. M., Martin, G. B., McNatty, K. P., and McNeilly, A. S. (1993). A model for follicle selection and the determination of ovulation rate in the ewe. *Reproduction, Fertility, and Development*, 5(5):459–78.

Scaramuzzi, R. J., Baird, D. T., Campbell, B. K., Driancourt, M.-A., Dupont, J., Fortune, J. E., Gilchrist, R. B., Martin, G. B., McNatty, K. P., McNeilly, A. S., Monget, P., Monniaux, D., Viñoles, C., and Webb, R. (2011). Regulation of folliculogenesis and the determination of ovulation rate in ruminants. *Reproduction, Fertility, and Development*, 23(3):444–67.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6):863–4.

Seitz, H., Royo, H., Bortolin, M.-L., Lin, S.-P., Ferguson-Smith, A. C., and Cavaillé, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Research*, 14(9):1741–8.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–504.

Sherratt, A. (1981). Plough and pastoralism: Aspects of the secondary products revolution. In *Pattern of the Past : Studies in Honour of David Clarke*, pages 261–305. Cambridge University Press, Cambridge [Eng.] ; New York.

Shi, J., Dong, M., Li, L., Liu, L., Luz-Madrigal, A., Tsonis, P. A., Del Rio-Tsonis, K., and Liang, C. (2015). mirPRo-a novel standalone program for differential expression and variation analysis of miRNAs. *Scientific Reports*, 5:14617.

Silva, B. D. M., Castro, E. A., Souza, C. J. H., Paiva, S. R., Sartori, R., Franco, M. M., Azevedo, H. C., Silva, T. A. S. N., Vieira, A. M. C., Neves, J. P., and Melo, E. O. (2011). A new polymorphism in the Growth and Differentiation Factor 9 (GDF9) gene is associated with increased ovulation rate and prolificacy in homozygous sheep. *Animal Genetics*, 42(1):89–92.

Silva, C. L. A. D., van den Brand, H., Laurenssen, B. F. A., Broekhuijse, M. L. W. J., Knol, E. F., Kemp, B., and Soede, N. M. (2016). Relationships between ovulation rate and embryonic and placental characteristics in multiparous sows at 35 days of pregnancy. *Animal*, 10(7):1192–9.

Sistiaga-Poveda, M. and Jugo, B. M. (2014). Evolutionary dynamics of endogenous Jaagsiekte sheep retroviruses proliferation in the domestic sheep, mouflon and Pyrenean chamois. *Heredity*, 112(6):571–8.

Soneson, C., Love, M. I., and Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*, 4:1521.

Song, G., Spencer, T. E., and Bazer, F. W. (2006). Progesterone and interferon-tau regulate cystatin C in the endometrium. *Endocrinology*, 147(7):3478–83.

Sormunen-Cristian, R. and Jauhiainen, L. (2002). Effect of nutritional flushing on the productivity of Finnish Landrace ewes. *Small Ruminant Research*, 43(1):75–83.

Souza, C. J. H., McNeilly, A. S., Benavides, M. V., Melo, E. O., and Moraes, J. C. F. (2014). Mutation in the protease cleavage site of GDF9 increases ovulation rate and litter size in heterozygous ewes and causes infertility in homozygous ewes. *Animal Genetics*, 45(5):732–9.

Spencer, T. E., Forde, N., and Lonergan, P. (2016). The role of progesterone and conceptus-derived factors in uterine biology during early pregnancy in ruminants. *Journal of Dairy Science*, 99(7):5941–50.

Spencer, T. E. and Hansen, T. R. (2015). Implantation and establishment of pregnancy in ruminants. *Advances in Anatomy, Embryology, and Cell Biology*, 216:105–35.

Spencer, T. E., Johnson, G. A., Bazer, F. W., and Burghardt, R. C. (2004a). Implantation mechanisms: Insights from the sheep. *Reproduction*, 128(6):657–68.

Spencer, T. E., Johnson, G. A., Bazer, F. W., Burghardt, R. C., and Palmarini, M. (2007). Pregnancy recognition and conceptus implantation in domestic ruminants: Roles of progesterone, interferons and endogenous retroviruses. *Reproduction, Fertility, and Development*, 19(1):65–78.

Spencer, T. E., Johnson, G. A., Burghardt, R. C., and Bazer, F. W. (2004b). Progesterone and placental hormone actions on the uterus: Insights from domestic animals. *Biology of Reproduction*, 71(1):2–10.

Spencer, T. E. and Palmarini, M. (2012). Endogenous retroviruses of sheep: A model system for understanding physiological adaptation to an evolving rumi-

nant genome. *Journal of Reproduction and Development*, 58(1):33–7.

Spencer, T. E., Sandra, O., and Wolf, E. (2008). Genes involved in conceptus–endometrial interactions in ruminants: Insights from reductionism and thoughts on holistic approaches. *Reproduction*, 135(2):165–79.

Srivastava, A., Malik, L., Sarkar, H., Zakeri, M., Soneson, C., Love, M. I., Kingsford, C., and Patro, R. (2020). Alignment and mapping methodology influence transcript abundance estimation. *Genome Biology*, 21:239.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics*, 20(11):631–56.

Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Tosser-Klopp, G., and Arranz, J. J. (2017). Variant discovery in the sheep milk transcriptome using RNA sequencing. *BMC Genomics*, 18:170.

Sun, Z., Evans, J., Bhagwate, A., Middha, S., Bockol, M., Yan, H., and Kocher, J.-P. (2014). CAP-miRSeq: A comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics*, 15:423.

Thieringer, H. A., Jones, P. G., and Inouye, M. (1998). Cold shock and adaptation. *Bioessays*, 20(1):49–57.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2):178–92.

Tixier-Boichard, M., Verrier, E., Rognon, X., and Zerjal, T. (2015). Farm animal genetic and genomic resources from an agroecological perspective. *Frontiers in Genetics*, 6:153.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5):511–5.

Usman, T., Hadlich, F., Demasius, W., Weikard, R., and Kühn, C. (2017). Unmapped reads from cattle RNAseq data: A source for missing and misassembled sequences in the reference assemblies and for detection of pathogens in the host. *Genomics*, 109(1):36–42.

Våge, D. I., Husdal, M., Kent, M. P., Klemetsdal, G., and Boman, I. A. (2013). A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. *BMC Genetics*, 14:1.

van Karnebeek, C. D., Sly, W. S., Ross, C. J., Salvarinova, R., Yaplito-Lee, J., Santra, S., Shyr, C., Horvath, G. A., Eydoux, P., Lehman, A. M., Bernard, V., Newlove, T., Ukpeh, H., Chakrapani, A., Preece, M. A., Ball, S., Pitt, J., Vallance, H. D., Coulter-Mackie, M., Nguyen, H., Zhang, L.-H., Bhavsar, A. P., Sinclair, G., Waheed, A., Wasserman, W. W., and Stockler-Ipsiroglu, S. (2014). Mitochondrial carbonic anhydrase VA deficiency resulting from CA5A alterations presents with hyperammonemia in early childhood. *American Journal of Human Genetics*, 94(3):453–61.

Vegetti, W. and Alagna, F. (2006). FSH and folliculogenesis: From physiology to ovarian stimulation. *Reproductive BioMedicine Online*, 12(6):684–94.

Wang, W., Zhang, X., Zhou, X., Zhang, Y., La, Y., Zhang, Y., Li, C., Zhao, Y., Li, F., Liu, B., and Jiang, Z. (2019). Deep genome resequencing reveals artificial and natural selection for visual deterioration, plateau adaptability and high prolificacy in Chinese domestic sheep. *Frontiers in Genetics*, 10:300.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.

Weldenegodguad, M., Pokharel, K., Ming, Y., Honkatukia, M., Peippo, J., Reilas, T., Røed, K. H., and Kantanen, J. (2020). Genome sequence and comparative analysis of reindeer (Rangifer tarandus) in northern Eurasia. *Scientific Reports*, 10:8980.

Whitacre, L. K., Tizioto, P. C., Kim, J., Sonstegard, T. S., Schroeder, S. G., Alexander, L. J., Medrano, J. F., Schnabel, R. D., Taylor, J. F., and Decker, J. E. (2015). What's in your next-generation sequence data? An exploration of unmapped DNA and RNA sequence reads from the bovine reference individual. *BMC Genomics*, 16:1114.

Xu, S.-S., Gao, L., Xie, X.-L., Ren, Y.-L., Shen, Z.-Q., Wang, F., Shen, M., Eyorsdottir, E., Hallsson, J. H., Kiseleva, T., Kantanen, J., and Li, M.-H. (2018). Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. *Frontiers in Genetics*, 9:118.

Yang, H., Wang, F., Li, F., Ren, C., Pang, J., Wan, Y., Wang, Z., Feng, X., and Zhang, Y. (2018). Comprehensive analysis of long noncoding RNA and mRNA expression patterns in sheep testicular maturation. *Biology of Reproduction*, 99(3):650–61.

Ying, S. J., Xiao, S. H., Wang, C. L., Zhong, B. S., Zhang, G. M., Wang, Z. Y., He, D. Y., Ding, X. L., Xing, H. J., and Wang, F. (2013). Effect of nutrition on plasma lipid profile and mRNA levels of ovarian genes involved in steroid hormone synthesis in Hu sheep during luteal phase. *Journal of Animal Science*, 91(11):5229–39.

Yurchenko, A. A., Daetwyler, H. D., Yudin, N., Schnabel, R. D., Jagt, C. J. V., Soloshenko, V., Lhasaranov, B., Popov, R., Taylor, J. F., and Larkin, D. M. (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Scientific Reports*, 8:12984.

Zhang, Y., Li, F., Feng, X., Yang, H., Zhu, A., Pang, J., Han, L., Zhang, T., Yao, X., and Wang, F. (2017). Genome-wide analysis of DNA methylation profiles on sheep ovaries associated with prolificacy using whole-genome Bisulfite sequencing. *BMC Genomics*, 18:759.

Zhang, Z., Tang, J., Di, R., Liu, Q., Wang, X., Gan, S., Zhang, X., Zhang, J., Hu, W., and Chu, M. (2019). Comparative transcriptomics reveal key sheep (Ovis aries) hypothalamus LncRNAs that affect reproduction. *Animals*, 9(4):152.